J. Clin. Chem. Clin. Biochem.
Vol. 18, 1980, pp. 461–490

Determination of Assigned Values in Control Specimens for Internal Accuracy Control and for Interlaboratory Surveys

Evaluation of 200 Different Lots with Identical Experimental Design: Experiences and Conclusions

By *E. Hansert* and *D. Stamm*

Department of Biostatistics and Department of Clinical Chemistry, Max-Planck-Institut für Psychiatrie, Munich

(Received August 14, 1979/March 27, 1980)

In honor of Professor Dr. Dr. h. c. Peter Karlson on the occasion of his 60th birthday

Summary: In the Federal Republic of Germany there are guidelines for a basic program for internal and external quality control of quantitative clinical chemical analyses prepared by the German medical society in connection with the Calibration Act. These guidelines specify that for internal accuracy control and for interlaboratory surveys, control specimens are to be used with assigned values and assigned intervals that have been determined by particularly well-qualified, independent laboratories serving as reference laboratories. The assigned values are often different from the best estimate (by measurement) of the "true value." They are influenced by the matrix of the specimen and by the method used, and they must be determined under routine conditions.

The protocol for obtaining the analytical results under carefully defined conditions is described in detail. The statistical treatment of the analytical results and the authors' observations on the structure of the data and the properties of the dispersion and location parameters for the individual laboratories are discussed. These observations led to the development of a procedure for selecting assigned intervals and assigned values in which no special assumptions are made about how the data are distributed. This procedure is described in detail.

The importance of the authors' observations for the evaluation of analytical results in the clinical laboratory and for the statistical treatment of such data is discussed.

Comparative studies were carried out using the same control specimens with other designs for assigned value determination (e. g. consensus value principle). The advantages of the design presented here are discussed.

Sollwert-Ermittlung in Kontrollproben für die laborinterne Richtigkeitskontrolle und für Ringversuche: Erfahrungen an 200 Chargen mit derselben Versuchsanordnung

Zusammenfassung: In der Bundesrepublik Deutschland gibt es im Zusammenhang mit dem Eichgesetz Richtlinien der Bundesärztekammer für ein Basisprogramm zur internen und externen Qualitätskontrolle von quantitativen klinisch-chemischen Untersuchungen. Für die laborinterne Richtigkeitskontrolle und die Ringversuche werden Kontrollproben benutzt, deren Sollwerte und Sollwertbereiche durch besonders qualifizierte unabhängige Referenzlaboratorien ermittelt werden müssen. Die Sollwerte unterscheiden sich häufig von der besten Annäherung an den „wahren Wert". Sie sind von der Matrix der Probe und der benutzten Methode abhängig und müssen unter Routinebedingungen ermittelt werden.

Die Gewinnung der Analysenergebnisse unter den genau definierten Bedingungen wird mit allen Einzelheiten des Versuchsplanes beschrieben. Die statistische Behandlung der Analysenergebnisse, ihre Datenstruktur, ihre Streuungskriterien, ihre Lagekriterien in den Einzellaboratorien werden anhand von Beispielen mitgeteilt. Daraus wird das nicht von speziellen Verteilungsannahmen ausgehende Verfahren zur Festlegung der Sollbereiche und des Sollwertes hergeleitet.

Die Bedeutung dieser Erfahrungen für die Beurteilung der Zuverlässigkeit und Vergleichbarkeit quantitativer klinisch-chemischer Analysenergebnisse und die statistische Behandlung solcher Daten wird diskutiert.

Der Vergleich mit anderen Versuchsplänen zur Sollwert-Ermittlung (wie z. B. dem Consensus Value) erfolgt durch Anwendung beider Verfahren an denselben Kontrollproben. Die Vorzüge des Verfahrens werden diskutiert.

0340-076X/80/0018-0461$2.00
© by Walter de Gruyter & Co. · Berlin · New York

1. Introduction

In the ongoing control of clinical laboratory analysis for systematic errors, control specimens are used in which the concentration of the analyte, approximated by the assigned value, is known as exactly as possible (1–7).

For the monitoring process to be effective, the composition of the control specimens must be as close to that of the patient specimens as possible. Put another way, the matrix of the control specimen and that of the patient specimen should be very similar (8). By matrix is meant the sum of the major and minor components and their structures, in which the analyte is embedded.

As is well-known, many analytical methods used in clinical chemistry are not completely specific. Thus the analytical results may include not only the component "analyte" but also, depending on the matrix, nonspecific components of varying sizes. In addition, procedural deficiencies may be present as positive or negative components. Consequently, analytical results from primary standards, control specimens and patient specimens will usually have quite different components (Table 1-1).

If different methods are used to assay a particular constituent, this may lead to quite different results, as is evident if one looks at assigned values taken from the package inserts of commercially available control specimens (Table 1-2). This is an indication that the nonspecific components and the procedural deficiencies are different for different methods. They can contribute as much as 25% of the best estimate (by measurement) of the "true value" of the analyte. Furthermore, the nonspecific components can also vary considerably from lot to lot of the same control specimen.

Thus the question arises of whether for the ongoing control of systematic errors in routine clinical chemical analyses it is sufficient to know the best estimate of the "true value" of the analyte. In over a decade of involvement in internal quality control and interlaboratory surveys it has become clear that it is more effective to monitor results by comparing the analytical results for a control specimen with an assigned value that is dependent on the method used—and thus includes all of the nonspecific components and procedural deficiencies of that method—than by comparing the results with the best estimate of the "true value" of the analyte.

Tab. 1-1. Components of analytical results.

Solutions (Index)	Reading	Components				
		Best estimate of "true value"	Deficiencies in procedure	Nonspecific component	Interference component	Result, e. g. concentration
	(E)	(T)	(D)	(N)	(I)	
Primary standard solution (S)	E_S	c_{TS}	$\pm c_{DS}$			$= c_S$
Control specimen (C) (Standard specimen)	E_C	c_{TC}	$\pm c_{DC}$	$\pm c_{NC}$		$= c_C$
Patient specimen (P)	E_P	c_{TP}	$\pm c_{DP}$	$\pm c_{NP}$	$\pm c_{IP}$	$= c_P$

Tab. 1-2. Comparison of assigned values using different methods.

Constituent (unit) Method	Assigned value (Assigned interval)	
Creatinine (mg/100 ml)	Reg. No. 16900 lyophilized	Reg. No. 17100 lyophilized
Jaffé reaction		
– after precipitation of the protein and specific adsorption on fuller's earth	1.12 (1.01–1.23)	1.65 (1.51–1.79)
– kinetic	0.95 (0.85–1.05)	1.5 (1.2 –1.7)
– AutoAnalyzer	1.2 (1.1 –1.3)	1.7 (1.5 –1.9)
– after precipitation of the protein with trichloroacetic acid	1.39 (1.27–1.51)	1.94 (1.77–2.11)
Total protein (g/l)	Reg. No. 16900 lyophilized	Reg. No. 16600 liquid
Kjeldahl method	51.5 (50.5–52.5)	–
Biuret reaction		
– without sample blank	56.0 (52.0–60.0)	55.0 (51.0–59.0)
– with sample blank	52.0 (48.0–56.0)	55.0 (51.0–59.0)

Our procedure for assigned value determination and our practical experience with this procedure will be discussed under two separate headings:

1. Analytical procedure and preliminary data processing:

This phase includes a discussion of the selection of reference laboratories, the qualifications required of reference laboratory heads, the selection of standards and methods, the protocoling of analytical results and preliminary statistical procedures.

2. Procedure for determining assigned values from the analytical results:

This phase includes a description of the sets of analytical results, a discussion of their statistical properties and the implications for the determination of assigned values and assigned intervals. Also included are a comparison of the locations of the distributions and the size of the standard deviations from the different laboratories (identical specimen and same analytical method), an evaluation of differences in location with different analytical methods, a discussion of certain mathematical problems and a comparison of different experimental designs for assigned value determination.

The findings reported here are based on analyses on more than 200 different lots of control specimens that were performed and processed according to the same experimental design. On the basis of these findings conclusions are drawn about the advantages and disadvantages of different procedures for assigned value determination for internal accuracy control and for the establishment of decision limits for interlaboratory surveys.

The extensive observations reported here illustrate the problems involved in the statistical processing of "hard" data from biological material. Thus, the significance of these observations clearly extends beyond the field of clinical chemistry.

2. Obtaining the Analytical Results

Reliable assigned values can be expected only if the analyses on which they are based have been performed in selected laboratories, chosen because they have superior facilities and are under highly qualified direction. Furthermore, the analytical methods must have been tested extensively and must be used under continuing control, with suitable standards and according to an appropriate, standardized protocol.

2.1 Selection of Reference Laboratories

In accordance with German law (9, 10), the associated Guidelines of the Medical Society of West Germany (Richtlinien der Bundesärztekammer) and the Implementation Regulations and Explanations (11, 12), the reference laboratories must be independent of the manufacturers and importers of equipment, reagents, standards and control specimens. The heads of the reference laboratories are personally responsible for the analytical results; they are selected by the medical societies of the Länder (countries) on the recommendation of the Medical Society of West Germany (MSWG) in cooperation with the scientific societies.

The reference laboratories must have available all such facilities and procedures as are necessary to insure the reliability of their analytical results. These are:

1. a comprehensive system of quality control,

2. facilities for testing the purity of substances that are used in the preparation of standards and also facilities for preparing standards,

3. facilities for comparison of methods according to Section 2.3.3 of the Guidelines,

4. ongoing external quality control by means of comparative studies with other reference laboratories.

The heads of the reference laboratories must be especially conversant with both the theoretical and practical aspects of this field and must be in a position to develop and test new methods themselves.

There should be only a limited number of reference laboratories.

According to the regulations for implementing the Guidelines of the MSWG (12), it is advantageous if several reference laboratories, designated as such by the medical societies of the Länder, form a group. This group conducts negotiations with the manufacturers of control specimens, organizes assigned value determinations, performs the statistical evaluation of the analytical results and computes the assigned values. It then transmits to the manufacturer or importer and to the participating reference laboratories the assigned values and assigned intervals and the various statistics for the individual laboratories based on their own analytical results only.

The Section on Assigned Value Determinations (headed by Prof. *D. Stamm*) of the Reference Commission of the German Society for Clinical Chemistry (GSCC, Deutsche Gesellschaft für Klinische Chemie) is such a body. Together with the Commission on Standardization of the GSCC and the head of the Department of Biostatistics of the Max Planck Institute for Psychiatry, Dr. *E. Hansert*, it has developed the procedure for assigned value determination described below. The procedure is carried out in accordance with an agreement between the reference laboratories and the manufacturers of control specimens.

This model for assigned value determination was tested in two long-term interlaboratory surveys, one involving laboratories in the Federal Republic of Germany only (13), the other involving the clinical chemistry societies in the Netherlands, Austria, Switzerland and the Federal Republic (14). Since the beginning of these surveys (1968), the same experimental design has been used for all assigned value determinations in the Federal Republic (15).

2.2 Selection of Standards and Methods

The analyses for assigned value determination must be performed with methods specified as Selected Routine Methods by the Commission on Standardization, where such methods are available. The standard materials specified must be used. In addition, assigned values are also determined with frequently used methods whose reliability is adequate for diagnostic purposes.

2.3 The Experimental Design

It must be possible to obtain from the analytical results both a location parameter (the assigned value) and a measure of variability to serve as the basis for internal accuracy control and for evaluation of interlaboratory surveys. At the time we developed our experimental design, there were, with the exception of a few pilot studies, no reliable data on the distribution of analytical results from different runs in the same laboratory, or on the pooling of results from different laboratories. For this reason a protocol had to be developed that would provide information on the distribution of the results within each laboratory. On the basis of this infor-

mation it would then be possible to decide whether to evaluate the data under the usual assumption of a normal distribution or whether other methods of evaluation would have to be sought.

In developing the protocol, it had to be taken into account that within a particular laboratory analyses were to be performed in different runs on different days. To be able to calculate the standard deviation from day to day, and in addition to get an idea of the size of the random error within a run, duplicate determinations were specified; the standard deviation of the series could then be calculated from the differences between the pairs of values. Since a new vial of the control specimen was to be opened for each run, the results would also provide information on variability from vial to vial (more of a problem some years ago than now) and on the homogeneity of the control specimen.

Also to be considered were certain restrictions on the design resulting from

1. the volume of control specimen material available,

2. the funds available for the analyses,

3. the number of laboratories qualified to be reference laboratories.

With these factors in mind the following protocol was set up:

1. Each participating reference laboratory makes duplicate determinations on 15 consecutive working days for certain constituent-method combinations. The analyses must be made under routine conditions along with analyses on patient specimens. If liquid control specimens are used, only 10 duplicate determinations are necessary.

2. If possible, at least three reference laboratories participate in the assigned value determination for each lot of a control specimen and for each constituent-method combination. This is the case about 90% of the time; for certain constituents and analytical methods of lesser importance only one or two laboratories may be involved.

2.4 Protocoling of the Analytical Results

Each reference laboratory records its analytical results on special protocol sheets coded with a registry number, the first three numerals of which identify the lot of the control specimen, the last two the reference laboratory (Fig. 2-1).

The laboratories are instructed to record their results unrounded, i. e. so that the last digit is an estimated digit and the second to last a significant digit. The total number of digits may vary from laboratory to laboratory. If considered advisable, information is also requested about the equipment and/or reagents used. For enzyme activities, the manufacturer and lot number are also requested. The head of the reference laboratory checks and signs all completed protocol sheets.

2.5 Computer Evaluation

After the protocol sheets have been returned, code numbers are entered for the constituent analyzed, the method used, the number of analytical results, the unit of measurement and the number of decimal places. Then all of the data are in a form suitable for transfer to punch cards.

The analytical results with each control specimen are processed in two steps.

2.5.1 The frequency distribution of the analytical results

For each analyte and each analytical method used the frequency distribution of the analytical results is determined for each reference laboratory (Fig. 2-2), with the first and second results of the duplicate determinations shown separately. At present these results are given in conventional units. For each distribution the computer printout also shows the mean ($\bar{x}$), standard deviation (s), coefficient of variation (V%) and $\bar{x} \pm 2s$.

The agreement between the duplicate determinations is shown as the "standard deviation of the series" (s*). This parameter indicates when the reliability of an individual determination should be checked.

A second computer printout (Fig. 2-3) shows the pooled data for all laboratories for a given constituent-method combination. The first and second results are shown separately to permit easier recognition of outliers and to enable the calculation of other statistics based on single determinations.

In addition the following are shown: the mean of the means ($\bar{\bar{x}}$), the mean of the standard deviations ($\bar{s}$), $\bar{\bar{x}} \pm 2\bar{s}$ and the median (m) of all results (the calculation and use of $\bar{s}$ goes back to the time when we assumed that the precision for a given method and specimen would be very similar in all of the laboratories). These statistics are given in both conventional and SI units.

The fact that the analytical results may have been recorded with different numbers of digits for the same method (see Section 2.4) is dealt with as follows. For all laboratories that used the same method, all of the analytical results are rounded to the smallest number of decimal places recorded. The width of the classes in the frequency distributions for both the individual and pooled distributions is the difference between two consecutive numbers with this number of decimal places. Under no circumstances is any further combination of classes made. However, the various statistics are always calculated from the unrounded results.

GERMAN SOCIETY FOR CLINICAL CHEMISTRY
REFERENCE COMMISSION
Section on Assigned Value Determinations

1	2	3	4	5
2	1	2	0	1

Control Specimen | Lab

Registry Number

PROTOCOL SHEET

Constituent:

6	7	8
	1	7

Method:

9	10	11
	1	1

To be filled out by lab

Protein
Biuret reaction with consideration of the sample blank

12	13	14	15	16	17	18	14	15	16	17	18
1	5	1	1	1	0	1	1	2	1	0	1

To be filled out by lab

Date	1st result (g/l)	2nd result (g/l)	Punch card column
23. 8. 77	64.	64.	20-24
24. 8. 77	64.	63.	25-29
26. 8. 77	62.	62.	30-34
29. 8. 77	64.	64.	35-39
30. 8. 77	62.	63.	40-44
31. 8. 77	63.	63.	45-49
1. 9. 77	66.	66.	50-54
2. 9. 77	62.	62.	55-59
5. 9. 77	62.	62.	60-64
6. 9. 77	62.	63.	65-69
New punch card Repeat columns 1-17	18 2	18 2	

To be filled out by lab

7. 9. 77	64.	67.	20-24
8. 9. 77	61.	63.	25-29
9. 9. 77	65.	66.	30-34
13. 9. 77	63.	64.	35-39
14. 9. 77	64.	64.	40-44

Date: 27. SEP. 1977

Signature: Büttner

Head of Reference Laboratory

Complete Address: (use rubber stamp if available)

Prof. Dr. med. et rer. nat.
J. Büttner
Medizinische Hochschule
Institut für Klinische Chemie

Fig. 2-1. Protocol sheet.

DEUTSCHE GESELLSCHAFT FUER KLINISCHE CHEMIE ** ZENTRALE REFERENZINSTITUTION ** ABT. FUER SOLLWERTERMITTLUNG

SOLLWERTERMITTLUNG IN DER RICHTIGKEITSKONTROLLPROBE

REGISTRIERNUMMER : 21200
WARENZEICHENNAME : KONTROLLOGEN-L
HERSTELLER :BEHRINGWERKE AG
CHARGE :3107
AUSDRUCKDATUM : 3. 11. 77

82. SEITE VON SEITEN

BESTANDTEIL 17 PROTEIN DIMENSION G/100ML

METHODE BIURET - REAKTION MIT PROBENLEERWERT

1. 01		6. 02		7. 03		15. 04		23. 05	
1.ERGEBNIS	2.ERGEBNIS	1.ERGEBNIS	2.ERGEBNIS	1.ERGEBNIS	2.ERGEBNIS	1.ERGEBNIS	2.ERGEBNIS	1.ERGEBNIS	2.ERGEBNIS
5.5 | 0	5.5 | 0	5.5 | 0	5.5 | 0	5.5 | 3	5.5 | 1	5.5 | 0	5.5 | 0	5.5 | 0	5.5 | 0
5.6 | 0	5.6 | 0	5.6 | 1	5.6 | 3	5.6 | 0	5.6 | 5	5.6 | 0	5.6 | 0	5.6 | 0	5.6 | 0
5.7 | 0	5.7 | 0	5.7 | 1	5.7 | 2	5.7 | 3	5.7 | 1	5.7 | 0	5.7 | 0	5.7 | 0	5.7 | 0
5.8 | 0	5.8 | 0	5.8 | 3	5.8 | 3	5.8 | 4	5.8 | 2	5.8 | 0	5.8 | 0	5.8 | 0	5.8 | 0
5.9 | 0	5.9 | 0	5.9 | 2	5.9 | 4	5.9 | 1	5.9 | 0	5.9 | 0	5.9 | 0	5.9 | 0	5.9 | 0
6.0 | 0	6.0 | 0	6.0 | 3	6.0 | 1	6.0 | 2	6.0 | 3	6.0 | 0	6.0 | 0	6.0 | 0	6.0 | 1
6.1 | 1	6.1 | 0	6.1 | 4	6.1 | 1	6.1 | 2	6.1 | 3	6.1 | 0	6.1 | 0	6.1 | 5	6.1 | 3
6.2 | 5	6.2 | 3	6.2 | 1	6.2 | 1	6.2 | 0	6.2 | 0	6.2 | 0	6.2 | 0	6.2 | 7	6.2 | 9
6.3 | 2	6.3 | 5	6.3 | 0	6.3 | 0	6.3 | 0	6.3 | 0	6.3 | 6	6.3 | 5	6.3 | 3	6.3 | 2
6.4 | 5	6.4 | 4	6.4 | 0	6.4 | 0	6.4 | 0	6.4 | 0	6.4 | 8	6.4 | 9	6.4 | 0	6.4 | 0
6.5 | 1	6.5 | 0	6.5 | 0	6.5 | 0	6.5 | 0	6.5 | 0	6.5 | 1	6.5 | 1	6.5 | 0	6.5 | 0
6.6 | 1	6.6 | 2	6.6 | 0	6.6 | 0	6.6 | 0	6.6 | 0	6.6 | 0	6.6 | 0	6.6 | 0	6.6 | 0
6.7 | 0	6.7 | 1	6.7 | 0	6.7 | 0	6.7 | 0	6.7 | 0	6.7 | 0	6.7 | 0	6.7 | 0	6.7 | 0
X = 6.3200/	6.3733	X = 5.9280/	5.8220	X = 5.8013/	5.8193	X = 6.3893/	6.3867	X = 6.1867/	6.1847
S = 0.1373/	0.1534	S = 0.1558/	0.1849	S = 0.2056/	0.2196	S = 0.0713/	0.0717	S = 0.0538/	0.0677
V = 2.1728/	2.4065	V = 2.6290/	3.1752	V = 3.5439/	3.7741	V = 1.1153/	1.1223	V = 0.8697/	1.0947
X-2S = 6.0454/	6.0666	X-2S = 5.6163/	5.4523	X-2S = 5.3901/	5.3801	X-2S = 6.2468/	6.2433	X-2S = 6.0791/	6.0493
X+2S = 6.5946/	6.6801	X+2S = 6.2397/	6.1917	X+2S = 6.2125/	6.2586	X+2S = 6.5319/	6.5300	X+2S = 6.2943/	6.3201
S* = 0.0775		S* = 0.1635		S* = 0.0786		S* = 0.0402		S* = 0.0416	
ANZAHL DER DATEN: 15		ANZAHL DER DATEN: 15		ANZAHL DER DATEN: 15		ANZAHL DER DATEN: 15		ANZAHL DER DATEN: 15	

DIESE DATEN SIND GEISTIGES EIGENTUM DER GESELLSCHAFT UND DUERFEN NUR MIT DEREN
AUSDRUECKLICHER SCHRIFTLICHER ZUSTIMMUNG BENUTZT UND VEROEFFENTLICHT WERDEN

Fig. 2-2. Computer printout: Frequency distributions of the analytical results shown for each laboratory separately (protein, biuret method).

DEUTSCHE GESELLSCHAFT FUER KLINISCHE CHEMIE ** ZENTRALE REFERENZINSTITUTION ** ABT. FUER SOLLWERTERMITTLUNG

SOLLWERTERMITTLUNG IN DER RICHTIGKEITSKONTROLLPROBE

REGISTRIERNUMMER : 21200
WARENZEICHENNAME : KONTROLLOGEN-L
HERSTELLER :BEHRINGWERKE AG
CHARGE :3107
AUSDRUCKDATUM : 3. 11. 77

83. SEITE VON SEITEN

BESTANDTEIL 17 PROTEIN DIMENSION G/100ML

METHODE BIURET - REAKTION MIT PROBENLEERWERT

50. 01		50. 02	
1.ERGEBNIS	2.ERGEBNIS	1.ERGEBNIS	2.ERGEBNIS
5.5| 0	5.5| 0	5.5| 0	5.5| 0
5.6| 0	5.6| 0	5.6| 0	5.6| 0
5.7| 0	5.7| 0	5.7| 0	5.7| 0
5.8| 1	5.8| 1	5.8| 1	5.8| 0
5.9| 3	5.9| 1	5.9| 7	5.9| 3
6.0| 6	6.0| 9	6.0| 2	6.0| 7
6.1| 4	6.1| 3	6.1| 4	6.1| 4
6.2| 1	6.2| 0	6.2| 0	6.2| 0
6.3| 0	6.3| 1	6.3| 1	6.3| 1
6.4| 0	6.4| 0	6.4| 0	6.4| 0
6.5| 0	6.5| 0	6.5| 0	6.5| 0
6.6| 0	6.6| 0	6.6| 0	6.6| 0
6.7| 0	6.7| 0	6.7| 0	6.7| 0

X	=	6.0113/ 6.0180	X	=	5.9853/ 6.0400
S	=	0.1032/ 0.1100	S	=	0.1203/ 0.1098
V	=	1.7172/ 1.8281	V	=	2.0097/ 1.8180
X-2S	=	5.8049/ 5.7980	X-2S	=	5.7448/ 5.8204
X+2S	=	6.2178/ 6.2380	X+2S	=	6.2259/ 6.2596
S*	=	0.0941	S*	=	0.0432
ANZAHL DER DATEN: 15			ANZAHL DER DATEN: 15		

DIESE DATEN SIND GEISTIGES EIGENTUM DER GESELLSCHAFT UND DUERFEN NUR MIT DEREN
AUSDRUECKLICHER SCHRIFTLICHER ZUSTIMMUNG BENUTZT UND VEROEFFENTLICHT WERDEN

Fig. 2-2. (cont.)

DEUTSCHE GESELLSCHAFT FUER KLINISCHE CHEMIE ** ZENTRALE REFERENZINSTITUTION ** ABT. FUER SOLLWERTERMITTLUNG

SOLLWERTERMITTLUNG IN DER RICHTIGKEITSKONTROLLPROBE

REGISTRIERNUMMER : 21200
WARENZEICHENNAME : KONTROLLOGEN-L
HERSTELLER :BEHRINGWERKE AG
CHARGE :3107
AUSDRUCKDATUM : 3. 11. 77

84. SEITE VON SEITEN

BESTANDTEIL 17 PROTEIN DIMENSION G/100ML

METHODE BIURET - REAKTION MIT PROBENLEERWERT

Z U S A M M E N F A S S U N G

1.ERGEBNIS	2.ERGEBNIS
5.5 | 3	5.5 | 1
5.6 | 1	5.6 | 8
5.7 | 4	5.7 | 3
5.8 | 9	5.8 | 6
5.9 | 13	5.9 | 8
6.0 | 13	6.0 | 21
6.1 | 20	6.1 | 14
6.2 | 14	6.2 | 13
6.3 | 12	6.3 | 14
6.4 | 13	6.4 | 13
6.5 | 2	6.5 | 1
6.6 | 1	6.6 | 2
6.7 | 0	6.7 | 1

	DIMENSION G/100ML	DIMENSION G/L
$\bar{X}$	6.0904	60.9042
S	0.1260	1.2603
$\bar{X}-2S$	5.8384	58.3836
$\bar{X}+2S$	6.3425	63.4248
MEDIAN	6.1000	61.0000

ANZAHL DER TEILNEHMER: 2

DIESE DATEN SIND GEISTIGES EIGENTUM DER GESELLSCHAFT UND DUERFEN NUR MIT DEREN AUSDRUECKLICHER SCHRIFTLICHER ZUSTIMMUNG BENUTZT UND VEROEFFENTLICHT WERDEN

Fig. 2-3. Computer printout: Frequency distribution of the analytical results for all laboratories combined.

DEUTSCHE GESELLSCHAFT FUER KLINISCHE CHEMIE ** ZENTRALE REFERENZINSTITUTION ** ABT. FUER SOLLWERTERMITTLUNG

SOLLWERTERMITTLUNG IN DER RICHTIGKEITSKONTROLLPROBE

REGISTRIERNUMMER : 21200
WARENZEICHENNAME : KONTROLLOGEN-L
HERSTELLER :BEHRINGWERKE AG
CHARGE :3107
AUSDRUCKDATUM : 3. 11. 77

BESTANDTEIL PROTEIN DIMENSION G/100ML

LABOR	METHODE	$\bar{x}$	S	$\bar{x}+2*S$	$\bar{x}-2*S$	V	S*
01	BIURET - REAKTION MIT PROBENLEERWERT	6.3200	0.1373	6.5946	6.0454	2.1728	0.0775
		6.3733	0.1534	6.6801	6.0666	2.4065	
03	BIURET - REAKTION OHNE PROBENLEERWERT	6.0920	0.1219	6.3358	5.8482	2.0010	0.0274
		6.0960	0.1078	6.3115	5.8805	1.7677	
06	BIURET - REAKTION OHNE PROBENLEERWERT	6.3853	0.1570	6.6994	6.0713	2.4591	0.1733
		6.2907	0.1695	6.6297	5.9516	2.6947	
06	BIURET - REAKTION MIT PROBENLEERWERT	5.9280	0.1558	6.2397	5.6163	2.6290	0.1635
		5.8220	0.1849	6.1917	5.4523	3.1752	
07	BIURET - REAKTION OHNE PROBENLEERWERT	6.3200	0.1292	6.5783	6.0617	2.0439	0.0519
		6.3447	0.1204	6.5855	6.1039	1.8978	
07	BIURET - REAKTION MIT PROBENLEERWERT	5.8013	0.2056	6.2125	5.3901	3.5439	0.0786
		5.8193	0.2196	6.2586	5.3801	3.7741	
15	BIURET - REAKTION MIT PROBENLEERWERT	6.3893	0.0713	6.5319	6.2468	1.1153	0.0402
		6.3867	0.0717	6.5300	6.2433	1.1223	
23	BIURET - REAKTION OHNE PROBENLEERWERT	6.5560	0.0639	6.6838	6.4282	0.9746	0.0452
		6.5567	0.0749	6.7065	6.4069	1.1423	
23	BIURET - REAKTION MIT PROBENLEERWERT	6.1867	0.0538	6.2943	6.0791	0.8697	0.0414
		6.1847	0.0677	6.3201	6.0493	1.0947	
50	BIURET - REAKTION OHNE PROBENLEERWERT	6.3100	0.1091	6.5282	6.0918	1.7288	0.1021
		6.3240	0.1138	6.5515	6.0965	1.7988	
50	BIURET - REAKTION OHNE PROBENLEERWERT	6.5093	0.1235	6.7563	6.2624	1.8971	0.0434
		6.5647	0.1139	6.7924	6.3369	1.7348	
50	BIURET - REAKTION MIT PROBENLEERWERT	6.0113	0.1032	6.2178	5.8049	1.7172	0.0941
		6.0180	0.1100	6.2380	5.7980	1.8281	
50	BIURET - REAKTION MIT PROBENLEERWERT	5.9853	0.1203	6.2259	5.7448	2.0097	0.0432
		6.0400	0.1098	6.2596	5.8204	1.8180	

Fig. 2-4. Computer printout: Statistics for one constituent by laboratory and method.

2.5.2 The individual statistics by constituents, method and laboratory

The statistics described above (Fig. 2-2) for each distribution individually are presented together in a separate printout (Fig. 2-4), so that for each analyte the following data are available: name of the laboratory head, analytical method used, $\bar{x}$, s, V%, $\bar{x} \pm 2s$ for the first and second results and s*. These data are given in both conventional and SI units (16) (two separate printouts). Again, all of the calculations are based on the unrounded analytical results.

A schematic representation of the experimental design including the data analysis described above is shown in Figure 2-5. The statistics just mentioned serve as the basis for further analysis.

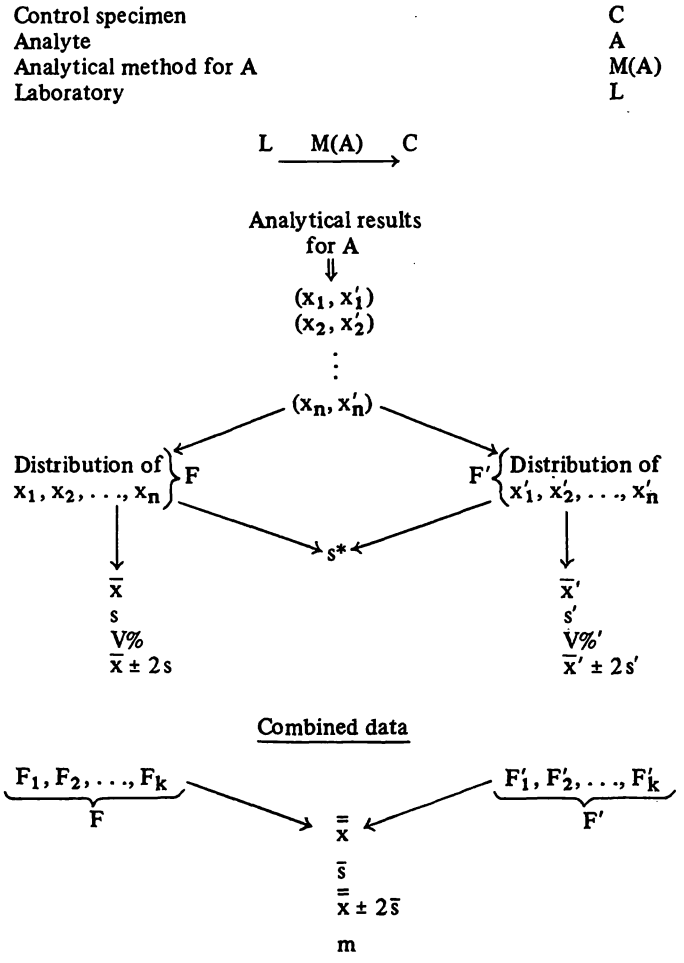

Fig. 2-5. Diagram of protocol and computer evaluation of the analytical results.

3. The Sets of Analytical Results: Characterization by Means of Assigned Values and Assigned Intervals

With the aid of statistical methods each set of analytical results obtained as described above is to be characterized by a location parameter (the assigned value) and a variation parameter or variation interval (confidence interval, assigned interval). Prior to describing the method we have developed we shall first discuss the purpose of assigned value determination from a statistical standpoint and describe the sets of analytical results obtained and their statistical properties.

3.1 The Purpose of Determining Assigned Values

For quality control in the clinical chemistry laboratory control specimens are needed whose composition with regard to certain substances can, for all practical purposes, be considered constant over a sufficiently long period of time. Numerical values that reflect this constant composition with adequate stability——sufficient stability to be used as reference values——are known as "assigned values" (Sollwerte). Assigned values represent sets of analytical results obtained by determination of the same constituent with the same method in several specimens of identical composition. The protocol used for replicate measurement depends on what needs to be taken into consideration in the assigned value.

For each analyte the following can be varied:

1. The "matrix" of the specimen: If the concentration of the analyte is held constant but the matrix is varied and if the analytical method is nonspecific this can have various effects.

2. The analytical method.

3. The participating laboratories: Different laboratories may perform their analyses under *equivalent* conditions, but these conditions will never be *identical.*

4. The number of determinations per laboratory, and the kind of determinations, e. g. single or duplicate.

The individual analytical result depends on the interplay of specimen, conditions of measurement and analytical method. An assigned value determination should permit recognition of the most important aspects of this interaction. Since a control specimen will normally be used in a particular laboratory for a relatively long period of time, it is advantageous if the protocol yields information on long-term aspects of the use of a particular analytical procedure with a particular specimen under identical conditions.

This means that it should be possible to evaluate both *repeatability* and *reproducibility* (the terms "repeatability" and "reproducibility" are used here in accordance with ISO/DIS 5725 (17): "Repeatability" refers to measurements in one laboratory over a short period of time and "reproducibility" to measurements over longer periods of time and/or from different laboratories).

Thus the ideal situation is one in which many laboratories perform many replicate analyses. However, for various reasons neither the number of participating laboratories per analytical method nor the number of analyses per laboratory can be chosen at will. It is probably the case that for each specimen and each analytical method a different combination, but one that is not known in advance, would be optimal.

In the following we shall discuss our experience using a protocol with 3 to 5 laboratories per analytical method and either 10 or 15 duplicate determinations per laboratory. This is not claimed to be a particularly good combination, but it provided sufficiently reliable information on certain aspects of the data that are not revealed by other protocols.

But first a number of formal observations on sets of analytical results.

3.2 Sets of Results: Their Description and Their Statistical Properties

From a formal point of view an assigned value is a value calculated with the aid of a set of analytical results. Since the analytical results are "random," an assigned value is also "random," and its statistical properties depend on the statistical properties of the set of results on which it is based.

A very important property that a set of results may have is that it is a *sample*. For this to be the case it is necessary that the associated random variables be identically and independently distributed (i. i. d. random variables). Thus a sample as used here is not one obtained by sampling from a given population. In the statistical literature on sets of results only one case is usually treated—either a true sample exists or any deviations from this property are assumed to be negligible. There is not even a name for sets of results other than samples. We therefore now introduce the term *conglomerate* ad hoc (with a *sample* being a special case of a conglomerate). A type of conglomerate seen frequently is shown in Figure 3-1. Since possible deviations from the property of a sample are assumed to be negligible, but are not usually investigated systematically, there are no practical criteria for determining when the property should no longer be assumed. We evaluated a number of sets of results to determine whether on the basis of simple criteria their members could be regarded as (significantly) different in the usual statistical sense. Some of our results are presented in Sections 3.4 and 3.5. The criterion is thus *statistical significance* of certain differences. Here we know, of course, that statistical significance does not necessarily mean *practical significance*; but the kinds of differences found also permit some observations on this point. In any case, evaluation of practical significance must be based on nonstatistical criteria and can only supplement the statistical evaluation.

The traditional, and still conventional, approach is to describe sets of results using the (arithmetic) mean as the location parameter and the standard deviation as the variation parameter (measure of precision, measure of dispersion). This concept is based on the model

$$x = \mu + e. \qquad (1)$$

In this model it is assumed that for a set of measurements of the same kind there is a "true value" and a

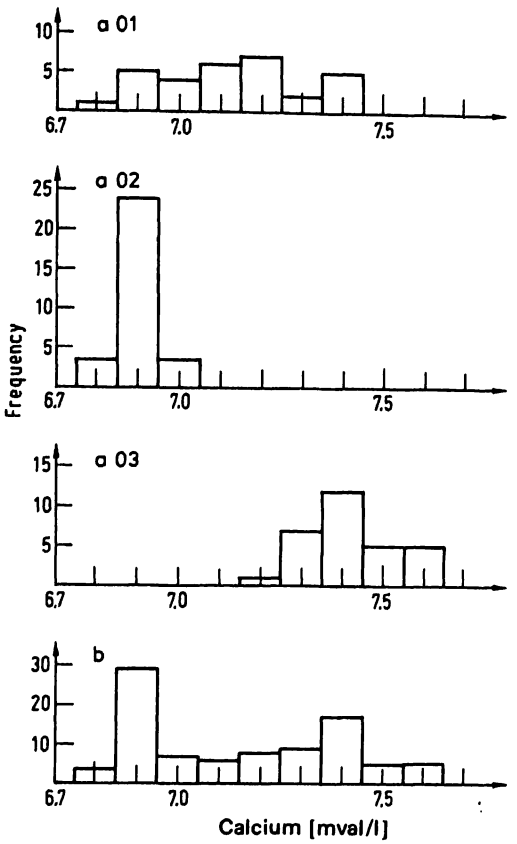

Fig. 3-1. Frequency distributions of the analytical results for calcium determination with atomic absorption spectrophotometry (Reg. No. 21500). Values given in the non-SI unit mval/l. For conversion into SI units (mmol/l) divide by 2.
a) Individual laboratories: a 01 Laboratory 01
a 02 Laboratory 02
a 03 Laboratory 03
b) All laboratories.

"measurement error"; the variability of the measurement is due to the variability of the measurement error, and "on the average" the measurement error disappears (otherwise it would be a "systematic error"). Therefore it is customary to regard a measurement as adequately described by μ together with the standard deviation $\sigma(e)$; for a sample of measurements, estimates of these parameters are the mean $\bar{x}$ and the standard deviation s. This model can be illustrated by a diagram (Fig. 3-2), where the abscissa represents the theoretical mean, that is, the expected value $E(x) = \mu$, and the ordinate the theoretical standard deviation $\sigma = \sigma(e)$.

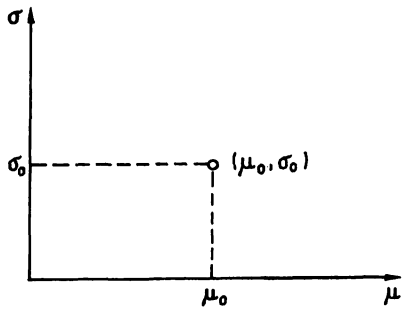

Fig. 3-2. Traditional measurement model for a single measurement $x = \mu + e$, where $\mu = \mu_0$, $\sigma(e) = \sigma_0$.

Model (1) is represented on this graph by a point (μ_0, σ_0) that is a particular value of (μ, σ).

Model (1) can be used whenever the set of results is a sample without any necessity of adopting the interpretation of μ as the "true value."

This model is complete only if the probability distribution of the measurements can be described completely by μ and σ(e).

This is true, for example, if x is distributed according to a normal distribution.

The first generalization of model (1) concerns analytical results from different laboratories. The traditional model for this case includes an additional component b for every laboratory L:

$$x = \mu + b_L + e \,. \qquad (2)$$

The additional assumptions are:

1. The laboratories are selected at random so that b itself is a random variable with a certain standard deviation σ(b); here, again, b is regarded as 0 "on the average"—because otherwise one would be dealing with a general systematic error.

2. The standard deviation σ(e) is independent of the laboratory.

3. The random variables b and e are stochastically independent.

In our (μ, σ) diagram, model (2) can be represented as in Figure 3-3. The line at σ_0 parallel to the abscissa represents the mathematically possible values for $\mu + b$, where the solid part of the line represents the realistically possible values and the dashed part is only of theoretical interest, e. g. if a normal distribution is assumed for b.

Model (2) is complete only if μ, σ(e) and σ(b) completely describe the distribution of x, especially when the additional assumptions 1 to 3 are actually the case.

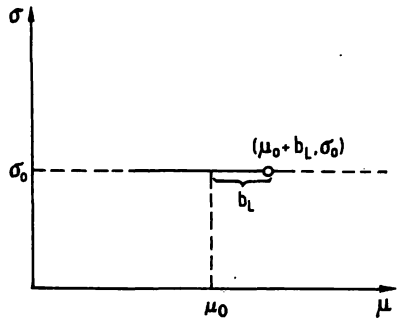

Fig. 3-3. Traditional measurement model for measurements $x = \mu + b_L + e_L$ from different laboratories, $\sigma(e_L) = \sigma_0$ for each laboratory.

Below and in Section 3.5 we will discuss why this model is also too idealized and does not suitably describe the sets of results actually occurring in assigned value determinations. In a more realistic model, σ(e) is also a variable. If it is a random variable, this would result in the diagram shown in figure 3-4. The shaded area represents a probability distribution for b and σ *together*. The equation for this model, somewhat simplified, could be as follows:

$$x = \mu + b_L + \sigma_L \cdot e_1 \,, \qquad (3)$$

where e_1 is a *standardized* random error, that is, it symbolizes a random error with a standard deviation of 1.

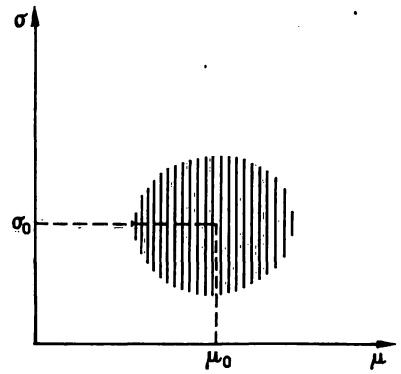

Fig. 3-4. Measurement model for measurements $x = \mu + e$, where μ and σ(e) are random variables.

But in the case of assigned value determination by reference laboratories it can hardly be assumed that the selection is random in the sense of the traditional model, and it is even less likely that there is a continuous probability distribution of b and that σ(e) is constant. It is more realistic to assume that there is a relatively small number of reference laboratories each with its own b and σ(e). This model can be represented by the following formula:

$$x = \mu_L + e_L \,, \qquad (4)$$

where

$$\mu_1, \mu_2, \ldots, \mu_L, \ldots$$
$$e_1, e_2, \ldots, e_L, \ldots$$
$$\sigma_1, \sigma_2, \ldots, \sigma_L, \ldots$$

are the expected values and the random errors of the laboratories actually participating ($\sigma_L = \sigma(e_L)$). In the (μ, σ) diagram there are then exactly the same number of dots as there are laboratories. Figure 3-5 gives an empirical example (Registry No. 21200, protein, biuret method taking the sample blank into account, six laboratories, one of which used two versions of the biuret method, indicated by ■). For sets of results with such properties the conclusion is that statistical evaluation using the traditional model is no longer justified, and a new approach must be developed.

3.3 The Selection of Assigned Values and Assigned Intervals

The observation is frequently made that analytical results from different laboratories, even if they are from samples of the same specimen and have been determined with the same analytical procedure, are not homogeneous with regard to either location or dispersion. Because of this, the problem of selecting an assigned value and the related assigned interval must be rethought. The usual definition of the assigned value as $\bar{x}$ and the assigned interval as $\bar{x} \pm 2s$ (the factor of 2 is related in a somewhat unclear manner to a probability of 95%—see Section 3.6) must be questioned for the following reasons:

1. It is assumed in these definitions (see Section 3.2) that the analytical results are distributed in an approximately symmetrical manner and, as far as the factor 2 is concerned, according to an approximately normal distribution. But these are assumptions that tend to be made because they fit well-known models rather than because they have a sound statistical basis.

2. Certain assumptions must be made about homogeneity. In particular, it must be assumed that the precision in all of the participating laboratories is the same. But as we found to our own surprise, this is not so.

3. Regarding allowable differences in location among the laboratories, in the above definitions the assumption must be made that the differences are themselves distributed according to an approximately normal distribution. Here again this assumption is made because it fits the conventional model and not—as far as we can tell—because it has been substantiated empirically. How kind we consider Nature to be when we assume that she always supplies us with just what we need to fit our conventional models!

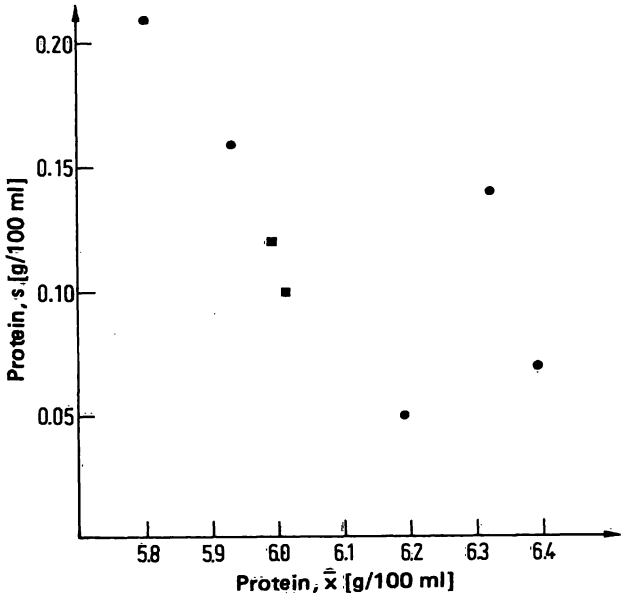

Fig. 3-5. Typical example of the location ($\bar{x}$) and imprecision (s) of results obtained in different laboratories under equivalent conditions (protein determination with the biuret reaction, Reg. No. 21200; 6 laboratories, one of which used two versions of the analytical method).

Comments

One sometimes reads in the literature that there is no need to be so careful about fulfilling the conditions for a normal distribution since symmetry, or sometimes merely the presence of a single modal value, is sufficient to guarantee that the statistical procedures suitable for normal distributions are accurate enough for the distribution under consideration. However, in the last two decades a great many mathematicians have been involved in a search for solutions to statistical problems that do not require the assumption of a normal distribution. And it is unlikely that they would have looked for alternatives if the normal distribution were so universally applicable.

In view of the situation that became apparent as we analyzed the participants' results in various assigned value determinations, and in the absence of a theory that takes such observations into account, an alternative had to be found that was practicable and at the same time did not contradict known theoretical principles. The objective was a description, with as few assumptions as possible, of sets of results from different laboratories where there are differences in location and dispersion and, in some cases, even in the kind of distribution.

Instead of constructing intervals on the basis of inapplicable theory the idea suggested itself of using "natural" constructions even though their theoretical properties would still have to be determined. Natural constructions are, for example, those which yield intervals including a certain specified minimum percentage of all observed values (such intervals would certainly be used more frequently if the belief in the normal distribution did not play such a dominant role in so many areas of statistical application). Such intervals—hereafter referred to as *P% intervals*—provide information about both location and dispersion. It would be worth considering whether or not P% intervals are the natural extension of the concept of measurement by means of a single value in those cases where the variability of measurement is so great that it cannot be disregarded in describing the results obtained.

It is clear that, because of the absence of assumptions, P% intervals would also be a useful way of describing conglomerates. The question remains as to which of the various possible kinds of construction should be used. In order to permit a discussion of the advantages and disadvantages of different constructions from a theoretical point of view, it would be desirable if appropriate theoretical studies were carried out.

A detailed discussion of the various alternatives is beyond the scope of this paper. Based on our past

experience (and on elementary theoretical considerations), we are currently using the following procedure:

1. The set of results to be described includes all first and second results from all participating laboratories (see also Comment 1, p. 475).

2. Let us first assume that no values appear more than once at the ends of the conglomerate (no tied values)—see Table 3-6, Example 1. It is advisable to have a total of at least 40 results (for the case where P = 95%). First the extreme values (the largest and smallest values in the conglomerate) are examined to determine which one is farther from the value adjacent to it (in Example 1, in which there are 90 results, this is the smallest value, 257). This value is omitted from the conglomerate, and the procedure is repeated with the remaining values until (ideally) exactly P% values remain. For reasons of tradition and because of the Guidelines ——not for theoretical reasons——the midpoint of the P% interval is designated as the assigned value. Thus in the case shown in Example 1 the values 257, 283, 287 and 296 are omitted (in that order); the resulting interval is actually a 95 5/9% interval, from 300 to 353, with assigned value 326.5. The procedure for omitting a value if in constructing the P% interval one has equivalent extreme values is described under 4. below.

3. In the ideal case (no tied values and decreasing distance between adjacent values as one moves inward), the P% interval resulting from this procedure is the smallest interval that includes the required number of values. In reality, however, tied values usually occur even at the extremes of the distribution——see Table 3-6, Example 2, which is the distribution on which the values in Example 1 are based. In the first step of constructing the P% interval we can proceed as in Example 1: The value 257 is omitted. In the second step 283 occurs twice, 353 only once. The procedure currently followed in such cases is to omit the value that occurs less frequently (here 353). If both values occur with the same frequency, the distance procedure is again applicable. The P% interval resulting in this way is often not the smallest possible interval, however.

4. If two equivalent values occur (i. e. two values with the same frequency and at the same distance from the next value), both are omitted if the P% interval will still contain enough values; otherwise the value is omitted that yields an interval midpoint closer to the median than if the other value were omitted. If the values are equivalent in this case, too, then a coin can be flipped or an additional point of reference (e. g. the mean) used.

5. An efficient computer algorithm can be devised to carry out this procedure automatically. However, we consider it better to proceed in such a way that it is possible to take into account any unusual features of the distribution. In a number of cases it seems appro-

Tab. 3-6. Examples of the procedure for selecting assigned intervals and assigned values. Values for chloride and potassium are the same in mmol/l as in mval/l

1. Simplified case: Lactate dehydrogenase determination (U/l), N = 90	
smallest values:	257-283-287-296-300
largest values:	353-351-347-346-345
Step 1: compare interval 257-283 with interval 353-351 omit: 257	
Step 2: compare interval 283-287 with interval 353-351 omit: 283	
Step 3: compare interval 287-296 with interval 353-351 omit: 287	
Step 4: compare interval 296-300 with interval 353-351 omit: 296	
Resulting interval, containing at least 95% of the results:	300-353
Assigned value:	326.5
2. Actual frequency distribution for above lactate dehydrogenase determination (U/l):	
smallest values (and frequency):	257(1)-283(2)-287
largest values (and frequency):	353(1)-351(3)-347
Step 1: compare:	257(1)-283 with 353(1)-351
omit:	257
Step 2: compare:	283(2)-287 with 353(1)-351
omit:	353
Step 3: compare:	283(2)-287 with 351(3)-347
omit:	283
Resulting interval, containing at least 95% of the results:	287-351
Assigned value:	319
3. Results from a chloride determination, coulometric method (mval/l), N = 90	
Results and frequency:	103 (1) 104 (1) 109 (10) 110 (13) 111 (21) 112 (18) 113 (21) 114 (3) 116 (2)
The usual algorithm yields:	interval 109-114 assigned value 111.5
Values actually selected:	interval 108-114 assigned value 111
4. Results from a potassium determination, Eppendorf flame photometry (mval/l), N = 90	
Results and frequency:	6.1 (1) 6.2 (18) 6.3 (44) 6.4 (7) 6.5 (10) 6.6 (4) 6.7 (6)
The usual algorithm yields:	interval 6.2-6.7 assigned value 6.45
Values actually selected:	assigned value 6.3 interval 5.9-6.7

priate to us to make changes in the resulting interval and/or to deviate from the procedure. For example:

a) The boundaries of the interval and the assigned value may be given in a greater number of figures than is actually useful in practice. Here the number can be reduced to a more practicable one (see Table 3-6, Example 3).

b) Near one or both ends of the set of results there may be such a concentration of values that some redistribution of these values seems necessary. The rule of thumb here is that if at least 10% of all results are concentrated at one end of the interval the interval is expanded at that end, the extent of this expansion depending on the structure of the set of results (see Table 3-6, Example 4).

c) On the other hand, there is sometimes such a concentration of results within a small interval that it seems advisable to designate a value from within this interval as the assigned value (possibly the mode) and then to determine the P% interval from here (see Table 3-6, Example 4).

Comments

1. If the procedure just described were carried out for the first and second results separately, the two assigned intervals would have to be combined subsequently. At present we consider it more advantageous to pool the first and second results and then to construct a 95% interval for the combined distribution. The objection can be raised that there could be a certain interdependence between the two results in a duplicate determination; but since it is only a matter of omitting extreme values and all other calculations are carried out for the first and second results separately, we feel that this interdependence can be ignored. The use of P% intervals as the basis for evaluating the results of participants in interlaboratory surveys is discussed elsewhere.

2. Whether it is always desirable to construct the narrowest possible P% interval is left open here. That was our original position, but in view of many distributions we have actually encountered, this appears at the moment to be too restrictive. Therefore we have modified the procedure in the manner described above to take into account the peculiarities of individual distributions (see 5. above). As already mentioned (p. 473), a discussion of the justification for such additional rules in terms of probability models is beyond the scope of this paper.

3. No matter how the interval is selected, there will always be cases in which the 95% interval includes more than 95% and even up to 100% of the values. The objection has been raised that in such cases it is absurd to continue to speak of 95% intervals. Therefore it is stated here explicitly that in the definition of P% intervals the value P is regarded as a (lower) limit, just as, for example, is the level of significance in certain statistical tests.

3.4 Differences in Location of the Results from Different Laboratories

A detailed discussion of the behavior of means and standard deviations of results from different laboratories is important because of the views that are widely held about these parameters. In particular, it is generally thought that a normal distribution can be assumed if replicate measurements are made under identical or equivalent conditions and that precision is the same if the same method of measurement is used under comparable conditions.

Of a whole complex of questions relating to distributions that will be dealt with in this paper, only one particular aspect will be discussed in this section. In a relatively large number of instances we have found that the distributions of results from several participants are almost completely or completely disjoint. Several examples are shown in Table 3-7. The examples have

Tab. 3-7. Differences in location of results from different laboratories. To obtain calcium concentrations in mmol/l, divide mval/l by 2.

Constituent Reg. No. (Method)	Range (Unit)	Reference laboratory (No. of results)						
Calcium								
Reg. No. 21400	(mval/l)	01	02	03				
(Atomic absorption spectrophotometry)	5.76–6.04	6	15	0				
	6.10–6.57	9	0	15				
Reg. No. 21500	(mval/l)	01	02	03				
	6.80–7.14	7	15	0				
	7.16–7.65	8	0	15				
Reg. No. 21500	(mval/l)	06	22	23				
(Emission flame photometry, Eppendorf flame photometry, acetylene flame)	6.10–7.15	7	15	0				
	7.18–7.38	8	0	15				
Serum iron								
Reg. No. 21500	(mg/100 ml)	02	06	14				
(Bathophenanthroline method with precipitation of protein)	134–147	10	15	0				
	148–179	5	0	15				
Creatinine								
Reg. No. 21000	(mg/100 ml)	06	30	50				
Jaffé reaction after precipitation of protein with trichloroacetic acid)	1.51–1.73	0	15	15				
	1.78–1.94	15	0	0				
Protein								
Reg. No. 21200	(g/100 ml)	01	06	07	15	23	50	50
(Biuret reaction with consideration of the sample blank)	5.5–6.0	0	10	13	0	0	10	10
	6.1	1	4	2	0	5	4	4
	6.2–6.7	14	1	0	15	10	1	1

been taken from four different determinations and, with the exception of protein, are presented as follows. The range including all results has been divided into a lower range and an upper range. The table shows the number of results per laboratory in each interval. In the first four examples the data have been divided in such a way that the two intervals contain as close to the same number of results as possible. In the creatinine example the data have been divided to show that one of the laboratories has a range that does not overlap at all with the range for the other two laboratories. In the last example, the number of results per laboratory that had the median value of 6.1 is also shown. Here again it is clear that the laboratories can be divided into two groups such that the results have little or no overlap. This example also shows that by pooling quite different distributions a new distribution may result that is almost symmetrical.

These examples reflect very clearly one of the points we want to stress again and again: The sets of results that can currently be obtained in assigned value determinations (considering the state of the art of analysis) are not as a rule homogeneous samples, but rather conglomerates consisting of several subsamples that may be quite different.

It is extremely unlikely that in a series of independent samples with n = 15 there will be two (or more) samples whose ranges have no overlap at all if the same distribution fits all of the samples. To be more specific, assuming the same distribution the number of such samples must be roughly 1000 if such an event is to have a relatively high probability of occurrence. This does not hold for a small number of determinations per laboratory and analytical method because the phenomenon just described then has a relatively high probability of occurrence even in the case of a common distribution (e. g. if there are 9 laboratories and 5 analyses per laboratory, yielding the same number of results as 3 laboratories and 15 analyses per laboratory, the probability would be about 25%).

A careful evaluation of the ranges in the assigned value determinations Reg. Nos. 21400 and 21500 for all analytical methods used by 3 participants (n = 15) showed at least 2 disjoint ranges for 7 of 35 and 9 of 37 analytical methods respectively; for Reg. No. 23200 the figures were even more striking: 15 of 31.

3.5 Differences in the Standard Deviations from Different Laboratories for the Identical Specimen and the Same Analytical Method

Whether one assumes a normal distribution or not and whether one considers the participating laboratories to be a random sample or not, the conventional statistical methods are based on model (1) or (2), where the standard deviation of e is assumed to be independent of the laboratories. At first we, too, made these assumptions in our data analysis, only to see with an increasing number of assigned value determinations that this precondition is not generally met and is probably the exception rather than the rule.

A simple statistical method for the comparison of several standard deviations based on the same number of analytical results is to divide the largest variance (s^2_{max}) by the smallest (s^2_{min}). Assuming a normal distribution, approximate critical values for the 5% and 1% levels can be found in the Biometrika Tables (18). With this method, suggested by *Hartley*, it is possible to get a quick impression of even a large number of sets of standard deviations. The quotient s^2_{max}/s^2_{min} gives a good picture by itself. The *Cochran* test (19), in which s^2_{max} is compared with $\Sigma\, s^2_L$, is preferred by many. The chief value of this test lies in identifying standard deviations that are too large. In our case, however, we must expect deviations in both directions, and for this reason we prefer to use the symmetric *Hartley* test.

Table 3-8 gives an overview of two parallel assigned value determinations (Reg. Nos. 21400 and 21500) for all constituent-method combinations where 3 laboratories participated. The value of the *Hartley* criterion for the standard deviations of the first results is shown. The critical value, assuming a normal distribution, is about 3.7 for the 5% level and about 5.3 for the 1% level. The cases where these values were exceeded are indicated by * and **, which does not mean, however, that these always correspond with the exact significance levels.

The data shown in Table 3-8 are typical of those evaluated thus far in assigned value determinations. Of course the exact number of significant variance quotients depends on the constituents actually analyzed and the analytical methods actually used and, in addition, on the specimens and on chance. There are constituents and methods where the standard deviations are relatively homogeneous but where for certain specimens great differences are suddenly seen that cannot be attributed to individual "outliers." This is the case, for example, in specimen with Reg. No. 21500 with the extremely high quotient of 70.9 for calcium determination with emission flame photometry, but no analytical grounds were found to justify the rejection of the analytical results. The same holds for the example shown in Table 3-7 with 7 sets of results for protein (Reg. No. 21200) with the biuret method, where the variance quotient of 14.5 far exceeds the 1% critical value (approximately 7.4). On the whole, no general predictions can be made about either the precision of an individual laboratory (for a specific analytical method) or the relationship of precision data from different laboratories to one another; rather, each specimen seems to have "a mind of its own." This is illustrated in Table 3-9 in terms of the coefficient of variation for creatinine, measured by the *Jaffé* reaction after adsorption on fuller's earth with precipitation of protein. The coefficient of variation was used here

Tab. 3-8. Quotient s^2_{max}/s^2_{min} for parallel assigned value determinations.
The two control specimens were analyzed in the reference laboratories in the same runs.

Constituent	Method	s^2_{max}/s^2_{min} for Reg. No. 21400	21500
Bilirubin	Photometry of the azopigment	18.9**	8.7**
Calcium	Atomic absorption spectrophotometry	13.4**	15.5**
	Emission flame photometry	4.8*	70.9**
Chloride	Coulometry	5.0*	7.3**
	Mercurimetric titration	5.8**	9.5**
Cholesterol	*Liebermann-Burchard* reaction	1.6	4.3*
	CHOD-PAP reaction	2.8	3.7
Glucose	Hexokinase reaction	3.6	2.2
	GOD-Perid	2.4	7.5**
	Glucose-dehydrogenase method	15.9**	11.3**
Uric acid	Destruction of uric acid by uricase with *Kagejama* reaction	17.5**	48.2**
Urea	Destruction of urea by urease with *Berthelot* reaction	5.3**	2.7
Potassium	Emission flame photometry with Li-guideline	7.0**	17.0**
	Emission flame photometry without Li-guideline	3.7	16.6**
Creatinine	*Jaffé* reaction after specific adsorption on fuller's earth	9.7**	3.4
	Jaffé reaction, kinetic	8.1**	5.2*
Copper	Photometry with bathocuproin	14.4**	4.8*
Magnesium	Atomic absorption spectrophotometry	6.7**	7.6**
Sodium	Emission flame photometry with Li-guideline	5.2*	7.5**
	Emission flame photometry without Li-guideline	10.0**	12.2**
Phosphorus	Reduction to molybdenum blue	2.2	19.6**
Protein	Biuret reaction without consideration of the sample blank	23.9**	32.5**
Creatine kinase	Standard method	11.0**	2.1
Aspartate aminotransferase	Standard method	3.1	2.6
Alanine aminotransferase	Standard method	1.1	1.8
Lactate dehydrogenase	Standard method	1.8	3.7
Alkaline phosphatase	Standard method	1.6	1.9
Amylase	Pharmacia	11.0**	3.7
Acid phosphatase	Total	4.6*	8.5**
	Tartrate	2.5	9.0**
Glutamate dehydrogenase	Standard method	2.2	3.2
Lithium	Atomic absorption spectrophotometry	1.1	7.6**
	Emission flame photometry	1.9	7.5**

* The critical value for the 5% level was exceeded.
** The critical value for the 1% level was exceeded.

because we can expect precision to depend to some extent on concentration. Table 3-9 shows the wide range of values both within and between laboratories.

3.6 Mathematical Considerations

This section contains an overview of those statistical facts and considerations that play a role in the treatment of sets of results.

3.6.1 Distributions

Without doubt the normal distribution plays a major part in mathematical statistics because of its formal properties, which permit exact treatment of a great number of problems, and because of its prominent position as a limiting distribution. No wonder one hopes that one's own data are distributed in a manner that can be treated so elegantly, thus enabling "textbook" analysis. One usually speaks of the assumption of a normal distribution, and that is exactly what it usually

Tab. 3-9. Dependence of precision on specimen.
Creatinine determination with the *Jaffé* reaction after precipitation of protein and specific adsorption on fuller's earth.

Control specimen Reg. No.	Assigned value (mg/100 ml)	Coefficients of variation (V%) of the reference laboratories 01	06	07	08
6000	1.9	5.3	1.3	2.6	
6100	1.95	5.8	1.6	3.0	
13500	1.35	3.2	2.9		4.5
13600	3.7	3.3	2.3		2.8
15500	1.65		2.9	3.2	4.1
15600	1.7		2.4	4.1	5.1
15700	3.55		3.7	3.6	5.2
16400	0.65		3.8		6.6
16500	0.90		4.0		4.6
16600	1.5		2.2		3.8
16700	1.3		1.6		3.2
21000	1.6	6.6	3.8	2.5	
21200	1.7	6.4	2.2		4.5
21400	1.5	5.1	1.8	2.5	
21500	4.0	2.9	1.9	3.6	

is. As long as there were not enough methods that could be used without this assumption and as long as the consequences of deviations from a normal distribution were not fully understood, the tendency to assume a normal distribution was understandable. But so many statistical questions now have solutions that require only minimal assumptions (referred to as nonparametric methods) that at least from this point of view there is usually no reason to persist in assuming a normal distribution.

In reality, no measuring procedure can yield a true normal distribution because the values that can be obtained with the procedure are necessarily limited to a relatively small range of measurement. Nevertheless, it is conceivable that the deviations from a normal distribution due to this factor only are so small that they can be ignored for all practical purposes. Furthermore, it is possible that in this practical sense an individual measurement may be distributed according to a normal distribution although the measurements taken as a whole are not. This is the case, for example, when both of the parameters (μ, σ) of the normal distribution to which the single measurement belongs can vary over time and when the times at which measurements are made are so far apart that this variability comes into play. This is often the case if measurements are made over a period of weeks or months——otherwise there would be no need to use control specimens in every run. If one assumes that such changes in the expected value and standard deviation also have a certain probability distribution $G(\mu, \sigma)$, then the interaction of the normal distribution of the individual measurement and the distribution G of the parameters of this normal distribution yields a "mixed distribution." This distribution can be expressed mathematically in terms of the probability density function of a normal distribution $\varphi(x/\mu, \sigma)$ and G:

$$\int \varphi(x/\mu, \sigma)\, dG(\mu, \sigma)\,, \qquad (5)$$

where the limits of the integral can vary from case to case. It is the exception, however, that this "mixture" produces a normal distribution. Another case where we can no longer speak of a normal distribution exists when the set of all results consists of subsets with different probability distributions; here the set of all results cannot be represented by a common distribution.

The reason often given for assuming a normal distribution is that the final result of an analysis is obtained by superimposing many components, and thus according to the limit theorems of probability calculus a random variable is obtained that for all practical purposes has a normal distribution, even if its components do not have such a distribution. But this is true only if in addition the following assumptions can be made:

1. The components are superimposed in an additive manner.

2. The degree of dependence among the components is sufficiently small.

3. There is no single component that dominates but does not itself have a normal distribution.

3.6.2 Statistical parameters

A normal distribution is unambiguously characterized by two parameters, the expected value μ and the variance σ^2 (or the standard deviation σ). These parameters are estimated on the basis of samples. An estimator is considered to be *optimal* if

1. it is unbiased, i. e. its expected value is identical to the parameter being estimated, and

2. the estimator has the minimum possible variance, i. e. it has maximum stability.

For normal distributions it is optimal to estimate the expected value μ with the mean of the sample $\bar{x}$ and nearly optimal to estimate σ^2 with s^2. But these properties may be lost if there are even small deviations from a normal distribution. In other words, these methods of estimation are not "robust." In recent years attempts have been made to investigate such phenomena systematically. An overview of this topic can be found, for example, in *Huber* (20). Thus, while $\bar{x}$ and s are particularly appropriate for normal distributions, it must be kept in mind that these parameters for characterizing a distribution are very special cases of location and dispersion parameters. In spite of the fact that $\bar{x}$ and s can be found even on pocket calculators, we must not forget that other parameters can be defined whose properties might make them more appropriate under certain circumstances.

For samples with variability or dispersion that is not negligible as compared with location, *intervals* are commonly used. The predominant use of $\bar{x}$ and s leads to the predominant use of interval descriptions such as

$$\bar{x} \pm c \cdot s\,, \qquad (6)$$

where a normal distribution is usually presupposed.

But the exact meaning of such a description usually remains unclear; one often hears the statement that a certain percentage of the analytical results lies within or can be expected to lie within such an interval. As a statement of fact this is unjustified, however, for the percentage of values within this interval is a random variable and in theory can have values between

$$100 \cdot \left(1 - \frac{1}{c^2}\right)\% \text{ and } 100\%$$

(assuming, of course, that $c > 1$). As a statistical statement it would relate to the probability distribution of this random percentage, a topic about which we refer the reader to *Albert & Johnson* (35).

Interval descriptions of type (6) can be interpreted as follows:

1. Confidence interval for the expected value μ: The constant c can be selected so that the interval (6) includes the true value of μ with a specified level of confidence.

2. Tolerance interval for a given percentage of the total probability: c can be selected so that the interval (6) includes at least a certain percentage of the total probability with a specified degree of certainty.

3. Tolerance interval for an additional analytical result: c can be selected so that the interval (6) will include an additional analytical result that is independent of the previous results with a specified probability.

It is clear that the value of c will depend to a large extent on the way the interval (6) is interpreted and also, of course, on the kind of probability distribution. For distributions that are not normal the first interpretation may not be meaningful since in this case the expected value may need to be replaced by a different location parameter. For the second interpretation there are nonparametric solutions. The third interpretation, like the first, loses its significance in all those instances where the mean is no longer sufficient to characterize the distribution.

As an alternative to intervals of type (6), whose usefulness depends too much on the assumptions made about the distribution, intervals can be selected in such a way that they contain *at least* a certain percentage of the actual analytical results. As already mentioned (Section 3.3), such intervals are referred to here as P% intervals. These intervals include both a description of location and, through their width, a description of dispersion. They can be interpreted as:

1. an estimate of a suitably defined P% interval of the underlying probability distribution,

2. a tolerance interval that includes a certain proportion of the total probability (depending on the manner of construction) with a specified degree of certainty.

3.6.3 Remarks on the outlier problem

Outliers are analytical results that appear to be too far away from the "other results." Preconditions for labeling a value as an outlier are (a) that the other results form a homogeneous sample and (b) that an assumption can be made about the kind of distribution. As might be expected, once again the normal distribution is the one most frequently assumed. If a value is identified as a statistical outlier, this does not automatically mean that the value is to be omitted (21). Rather, it means that a check must be made on the process through which the value was obtained and/or on whether the assumptions about the distribution type actually hold. The exclusion of extreme values on the basis of an assumed distribution only and without any other justification does lead to an improvement in the appearance of the distribution, but it may also be a case of circular reasoning. (It is something altogether different, of course, to exclude such values when calculating certain parameters.) This is also true if a laboratory submits several results all of which are classified as statistical outliers.

There are situations in which it is obvious that some error *must* have been made during the analytical process. In all cases where there has been any doubt (for whatever reason) it has been our practice to make inquiries at the laboratories involved and, if necessary, to request that the analyses be repeated. This has almost always led to a better understanding of the peculiarities of the particular analytical method *or* of the specimen; in many cases repeat analyses have yielded very similar extreme results. If such results were simply omitted, the remaining results would show greater homogeneity, but at the same time less would be learned from the analysis.

3.7 Differences in Location with Different Analytical Methods

For those analytical methods that are nonspecific the influence of different matrices cannot be predicted. Therefore, even when a constituent is analyzed in the same laboratory and in the same control specimen but with different methods, the amount of difference between methods cannot usually be predicted. The following example should make this clear. Table 3-10 shows values obtained in a single laboratory for the same constituent during several different assigned value determinations. The constituent was creatinine and the analytical methods used were: (1) *Jaffé* reaction after adsorption of fuller's earth with precipitation of protein, (2) *Jaffé* reaction after precipitation of protein with trichloroacetic acid and (3) *Jaffé* reaction, kinetic method. The means of the first results from 15 duplicate determinations by the method specified and the absolute and percent differences between methods 1 and 2 and methods 1 and 3 are shown. That the *size* of the differences depends on the specimen is evident.

3.8 Comparison of Different Experimental Designs

As pointed out earlier, in a protocol for determination of the composition of a certain specimen the following can be varied:

1. the number of participating laboratories,

2. the number of determinations per laboratory,

3. the kind of determination (single, replicate),

4. the time allowed for completion of the determinations,

5. the actual make-up of the participating laboratory team.

Tab. 3-10. Differences in location of results obtained with different analytical methods.

Creatinine determinations
carried out in parallel in reference laboratory 06 with the following methods:

1 *Jaffé* reaction after precipitation of protein and specific adsorption on fuller's earth
2 *Jaffé* reaction after precipitation of protein with trichloroacetic acid
3 *Jaffé* reaction, kinetic

Control specimen	Method	Results Mean	Differences between methods 1 and 2, 1 and 3	
		(mg/100 ml)	(mg/100 ml)	(%)
17100	1	1.53		
	2	1.75	.22	14.4
	3	1.28	.25	16.3
17600	1	1.46		
	2	1.65	.19	13.0
	3	1.21	.25	17.1
17700	1	3.42		
	2	3.47	.05	1.5
	3	2.63	.79	23.1
18700	1	1.08		
	2	1.21	.13	12.0
18800	1	1.31		
	2	1.45	.14	10.7
18900	1	1.31		
	2	1.55	.24	18.3
19000	1	3.80		
	2	4.01	.21	5.5
19200	1	1.63		
	2	1.78	.15	9.2
19300	1	1.79		
	2	1.99	.20	11.2
21400	1	1.40		
	2	1.75	.35	25.0
	3	1.36	.04	2.9
21500	1	4.08		
	2	4.41	.33	8.1
	3	3.44	.64	15.7

Before we review the available comparative data, some reflections are in order on fundamental aspects of experimental design. The following two extreme protocols will serve as our point of departure:

A. There are a large number of participating laboratories (e. g. all of those in a particular geographical or political region), and each makes a single determination.

B. A single laboratory makes a daily determination (single or replicate) over a relatively long period of time.

It is clear what information protocol B provides: adequate data on the location and precision of analytical results (and possibly even on the kind of distribution) obtained with a particular analytical method as used under the conditions prevailing in this laboratory. In addition, attention will be paid in this laboratory to difficulties in determining specific constituents with particular methods, and any difficulties will be reported to the person or commission in charge of the study. Furthermore, the laboratory will certainly be willing to make additional measurements under specially controlled conditions when such difficulties develop. On the other hand, no information is obtained about how much the location and precision of measurements on the specimen can vary if the measurements are made under conditions that are only equivalent rather than identical.

What information is obtained with protocols of type A? Let N be the number of laboratories that have used a particular method; in this instance N is also the number of analytical results obtained with the method. Now since for every laboratory L a characteristic location (expected value μ_L) and characteristic precision (standard deviation σ_L) can be expected (see Section 3.3), the following formulas hold:

$$E(\bar{x}) = \frac{1}{N} \Sigma \mu_L = \bar{\mu}_L , \qquad (7)$$

$$E(s^2) = \frac{1}{N} \Sigma \sigma_L^2 + \frac{1}{N-1} \Sigma (\mu_L - \bar{\mu}_L)^2 = \overline{\sigma_L^2} + s^2(\mu_L) . \qquad (8)$$

In words these formulas say: The expected value of the mean equals the mean of the expected values, the expected value of the variance equals the mean of the variances *plus* the variance of the means. Thus such a protocol does not provide any information on the contribution of variations within the individual laboratories or the contribution of differences in location between laboratories to the overall variability. Furthermore, such a protocol does not provide any information on whether laboratories with extremely poor precision (but possibly with a good central location) are among the participants. Laboratories with values lying at the extremes of the range of all values appear to be "bad" and those whose results are in the middle of the range appear to be "good," but whether precision is poor and to what extent the actual interaction of analytical method and specimen under identical conditions is reflected remains unclear.

If this kind of protocol is compared with one in which k laboratories provide n results per analytical method (so that $N = k \cdot n$), formulas (7) and (8) remain basically unchanged: $\bar{\mu}$ and σ_L^2 are still the mean expected value and mean variance for the k laboratories, and $s^2(\mu_L)$ is still the correct expression if each value μ_L is counted n times. This means that the value of $s^2(\mu_L)$ is generally smaller for $k < N$ laboratories than for N laboratories. Thus one would expect a larger dispersion for N values from N laboratories than for $k \cdot n$ values from k laboratories—assuming $\bar{\mu}_L$ and σ_L^2 to be identical in the two cases.

In the second case, however, σ_L^2 is a *variance from day to day*, while in the first case the *variance in the series*

should probably be used, which is known to be considerably smaller than the variance from day to day in some instances. So the two effects could conceivably balance each other out.

Twice we have had the opportunity to compare such a k · n protocol with a protocol of type A in two parallel assigned value determinations. Tables 3-11 to 3-14 show assigned values and assigned intervals for the two protocols to the extent that the analytical methods can be compared. For protocol (a) the assigned intervals were determined with $\bar{x} \pm 2s$, for the other protocol with the method described in Section 3.2. This limits comparability, however, because:

1. with one exception, k = 3 and n = 15 duplicate determinations were made for the k · n design, but for protocol A the total number of determinations N varied considerably from method to method.

2. protocol A was evaluated in two phases: First the mean and standard deviation were calculated for each constituent and all those values outside $\bar{x} \pm 3s$ eliminated. Then a new mean and standard deviation were calculated on the basis of the remaining values (this was done for the set of *all* results for each constituent; unfortunately the extent to which particular methods were involved could not be determined from the available data, but a marked reduction in the variance can generally be assumed).

Seen as a whole, the agreement of the two kinds of assigned values is astonishing. The intervals $\bar{x} \pm 2s$ in

Tab. 3-11. Comparison of results for assigned value determination obtained with two different protocols in the same control specimen (Reg. No. 17200). To obtain calcium values in mmol/l, divide mval/l by 2. Values for chloride are the same in mmol/l as in mval/l.

Constituent	(unit)	Method	Protocol *Whitehead* 1 analytical result per laboratory (Reg. No. X 1639)		Protocol Ger. Soc. Clin. Chem. 3 laboratories, 15 duplicate determinations each (Reg. No. 17200)
			No. of labs	Assigned value (assigned interval)	Assigned value (assigned interval)
Bilirubin	(mg/100 ml)	Photometry of the azopigment	35	1.2 (0.7 – 1.7)	1.1 (0.95– 1.25)
Calcium	(mval/l)	Atomic absorption spectrophotometry	54	5.7 (5.2 – 6.2)	5.8 (5.4 – 6.2)
Chloride	(mval/l)	Coulometry	66	102 (94.5 –109.5)	102 (100 –104)
		Mercurimetric titration	14	103 (96 –110)	105 (101 –109)
Cholesterol	(mg/100 ml)	*Liebermann-Burchard* reaction	35	94 (60 –128)	86 (76 – 96)
		CHOD-PAP reaction	26	85 (58 –112)	78 (72 – 84)
Serum iron	(µg/100 ml)	Bathophenanthroline method	29	163 (120 –206)	143 (126 –160)
Glucose	(mg/100 ml)	GOD-Perid	92	99 (91.5 –106.5)	99 (92 –106)
Uric acid	(mg/100 ml)	Destruction of uric acid by uricase with *Kagejama* reaction	20	4.6 (3.9 – 5.3)	4.6 (4.2 – 5.0)
		Phosphotungstic acid	39	5.0 (3.7 – 6.3)	4.65 (3.9 – 5.4)
Urea	(mg/100 ml)	Destruction of urea by urease with *Berthelot* reaction	49	40.5 (30 – 51)	40 (35 – 45)
Creatinine	(mg/100 ml)	*Jaffé* reaction after precipitation of protein with trichloroacetic acid	38	1.45 (0.85– 2.05)	1.5 (1.3 – 1.7)
Phosphorus	(mg/100 ml)	Reduction to molybdenum blue	73	3.25 (2.85– 3.65)	3.3 (3.0 – 3.6)
Protein	(g/100 ml)	Biuret reaction	77	6.8 (6.2 – 7.4)	6.7 (6.2 – 7.2)

protocol A are much wider in many cases than the assigned intervals obtained with the 3 X 15 protocol.

In order to investigate the properties of protocols of type k · n, the following procedure was used twice with several lots analyzed parallel to each other: Three laboratories made 15 duplicate determinations each with the most important analytical methods in accordance with the protocol already described. In addition, 2 laboratories performed 5 duplicate determinations each with each method. The following were evaluated for each method according to the procedure described in Section 3.3:

1. the results of the 3 participants with 15 duplicate determinations each,

2. the last 5 duplicate determinations of the 3 participants just mentioned together with the 5 duplicate determinations each from the 2 additional laboratories.

Even though the two groups of results were not obtained completely independently of one another, this procedure provided the opportunity to compare 3 X 15 designs with 5 X 5 designs. A detailed analysis will be reported in a later publication. All in all the assigned values and assigned intervals are extremely similar for the two designs (Table 3-15).

Tab. 3-12. Comparison of results for assigned value determination obtained with two different protocols in the same control specimen (Reg. No. 17300). To obtain calcium values in mmol/l, divide mval/l by 2. Values for chloride, given in mval/l correspond that in mmol/l.

Constituent	(unit)	Method	Protocol *Whitehead*		Protocol Ger. Soc. Clin. Chem.
			1 analytical result per laboratory (Reg. No. X 2739)		3 laboratories, 15 duplicate determinations each (Reg. No. 17300)
			No. of labs	Assigned value (assigned interval)	Assigned value (assigned interval)
Bilirubin	(mg/100 ml)	Photometry of the azopigment	43	3.4 (2.7 – 4.1)	3.35 (3.1 – 3.6)
Calcium	(mval/l)	Atomic absorption spectrophotometry	49	7.0 (6.5 – 7.5)	7.1 (6.6 – 7.6)
Chloride	(mval/l)	Coulometry	75	118 (112 –124)	116 (113 –119)
		Mercurimetric titration	17	118 (110 –126)	122 (117 –127)
Cholesterol	(mg/100 ml)	*Liebermann-Burchard* reaction	40	112 (79 –145)	106 (97 –115)
		CHOD-PAP reaction	29	94 (76 –112)	89 (82 – 96)
Serum iron	(µg/100 ml)	Bathophenanthroline method	30	174 (94 –254)	158 (138 –178)
Glucose	(mg/100 ml)	GOD-Perid	88	234 (217 –251)	234 (221 –247)
Uric acid	(mg/100 ml)	Destruction of uric acid by uricase with *Kagejama* reaction	18	8.8 (6.9 – 10.7)	8.4 (7.4 – 9.4)
		Phosphotungstic acid	40	9.4 (7.0 – 11.8)	9.2 (8.1 – 10.3)
Urea	(mg/100 ml)	Destruction of urea by urease with *Berthelot* reaction	54	94 (80 –108)	93 (85 –101)
Creatinine	(mg/100 ml)	*Jaffé* reaction after precipitation of protein with trichloroacetic acid	48	4.35 (3.15– 5.55)	4.5 (4.2 – 4.8)*
Phosphorus	(mg/100 ml)	Reduction to molybdenum blue	67	5.65 (5.0 – 6.3)	5.7 (5.2 – 6.2)
Protein	(g/100 ml)	Biuret reaction	78	6.95 (6.45– 7.45)	7.0 (6.6 – 7.4)

* only two participant laboratories

4. Discussion

For internal control of accuracy and for interlaboratory surveys according to the Guidelines of the Medical Society of West Germany (MSWG) (11, 12) suitable control specimens are needed in which the concentration of the analytes has been determined as reliably as possible (assigned values). In order to evaluate the interlaboratory surveys it is also necessary to have a measure of variation for the analytical results from the reference laboratories.

4.1 Determination of Analytical Results under Carefully Defined Routine Conditions

Because of the differences in specificity of different routine methods, assigned values depend to a greater or lesser degree on the method used (see Section 3.7). Therefore they must be determined for each of the most reliable and most frequently used routine methods and in each individual lot of a control specimen.

The assigned values are determined with statistical methods on the basis of analytical results from well-

Tab. 3-13. Comparison of results for assigned value determination obtained with two different protocols in the same control specimen (Reg. No. 21400). To obtain calcium values in mmol/l, divide mval/l by 2. Values for chloride, sodium and potassium are the same in mmol/l as in mval/l.

Constituent	(unit)	Method	Protocol *Whitehead* 1 analytical result per laboratory (Reg. No. S 0836)		Protocol Ger. Soc. Clin. Chem. 3 laboratories, 15 duplicate determinations each (Reg. No. 21400)
			No. of labs	Assigned value (assigned interval)	Assigned value (assigned interval)
Bilirubin	(mg/100 ml)	Photometry of the azopigment	38	2.00 (1.40– 2.60)	2.05 (1.8 – 2.3)
Calcium	(mval/l)	Atomic absorption spectrophotometry	54	5.97 (5.53– 6.41)	6.1 (5.8 – 6.4)
Chloride	(mval/l)	Coulometry	72	103.6 (97.8 –109.4)	104.5 (102 –107)
		Mercurimetric titration	14	104.7 (97.4 –112.0)	109 (102 –116)
Cholesterol	(mg/100 ml)	*Liebermann-Burchard* reaction	29	104.3 (75.7 –132.9)	105 (94 –116)
		CHOD-PAP reaction	32	94.5 (67.1 –121.9)	92 (78 –106)
Serum iron	(µg/100 ml)	Bathophenanthroline method *without* precipitation of protein	28	192 (133 –251)	173 (138 –208)
		with precipitation of protein			177 (162 –192)
Glucose	(mg/100 ml)	GOD-Perid	94	98 (88 –108)	94 (88 –100)
Uric acid	(mg/100 ml)	Destruction of uric acid by uricase with *Kagejama* reaction	22	4.4 (3.2 – 5.6)	4.5 (4.1 – 4.9)
Urea	(mg/100 ml)	Destruction of urea by urease with *Berthelot* reaction	51	34.3 (27.2 – 41.4)	34 (31 – 37)
Potassium	(mval/l)	Emission flame photometry with Li-guideline	91	5.28 (5.09– 5.47)	5.25 (5.1 – 5.4)
Sodium	(mval/l)	Emission flame photometry with Li-guideline	93	145.5 (142 –149)	145 (142 –148)
Phosphorus	(mg/100 ml)	Reduction to molybdenum blue	62	2.78 (2.32– 3.24)	3.0 (2.7 – 3.3)
Protein	(g/100 ml)	Biuret reaction	68	6.73 (6.06– 7.40)	6.9 (6.5 – 7.3)

equipped reference laboratories under the direction of persons particularly well qualified for this task. The reference laboratories must be independent of the manufacturers of control specimens, standards, reagents and equipment (11, 12).

In the laboratories selected to serve as reference laboratories the control specimens are analyzed under carefully defined conditions; the standards and the methods to be used are specified to the greatest extent possible. The number of analyses to be made per run, the number of runs per working day and the total number of runs are also specified.

Since the assigned values determined in this way are used to monitor rountine methods, the analytical results on which they are based must have been obtained in runs of routine analyses along with, or as if they were, patient specimens. This is of particular importance with regard to dispersion; the dispersion parameter must be based on results obtained under routine conditions if it is to be of use in evaluating interlaboratory survey results.

Tab. 3-14. Comparison of results for assigned value determination obtained with two different protocols in the same control specimen (Reg. No. 21500). To obtain calcium values in mmol/l, divide mval/l by 2. Values for chloride, sodium and potassium are the same in mmol/l as in mval/l.

Constituent	(unit)	Method	Protocol *Whitehead* 1 analytical result per laboratory (Reg. No. S 1536)		Protocol Ger. Soc. Clin. Chem. 3 laboratories, 15 duplicate determinations each (Reg. No. 21500)
			No. of labs	Assigned value (assigned interval)	Assigned value (assigned interval)
Bilirubin	(mg/100 ml)	Photometry of the azopigment	43	3.42 (2.52– 4.32)	3.35 (3.0 – 3.7)
Calcium	(mval/l)	Atomic absorption spectrophotometry	47	7.16 (6.62– 7.70)	7.2 (6.8 – 7.6)
Chloride	(mval/l)	Coulometry	81	114.0 (107.1 –120.9)	114 (110 –118)
		Mercurimetric titration	17	118.4 (112.5 –124.3)	120 (114 –126)
Cholesterol	(mg/100 ml)	*Liebermann-Burchard* reaction	31	117.2 (88.0 –146.4)	112 (98 –126)
		CHOD-PAP reaction	39	101.0 (78.1 –123.9)	97 (88 –106)
Serum iron	(µg/100 ml)	Bathophenanthroline method *without* precipitation of protein	30	185.0 (129.5 –240.5	162 (134 –190)
		with precipitation of protein			153 (134 –172)
Glucose	(mg/100 ml)	GOD-Perid	106	231.7 (210.2 –253.2)	222 (209 –235)
Uric acid	(mg/100 ml)	Destruction of uric acid by uricase with *Kageyama* reaction	22	10.47 (8.45– 12.49)	10.4 (9.7 – 11.1)
Urea	(mg/100 ml)	Destruction of urea by urease with *Berthelot* reaction	59	89.6 (74.9 –104.3)	89 (82 – 96)
Potassium	(mval/l)	Emission flame photometry with Li-guideline	108	6.98 (6.75– 7.21)	7.0 (6.75– 7.25)
Sodium	(mval/l)	Emission flame photometry with Li-guideline	106	154.3 (150.3 –158.3)	154 (151 –157)
Phosphorus	(mg/100 ml)	Reduction to molybdenum blue	60	6.27 (5.77– 6.77)	6.4 (6.0 – 6.8)
Protein	(g/100 ml)	Biuret reaction	59	7.20 (6.80– 7.60)	7.4 (7.0 – 7.8)

Tab. 3-15. Comparison of assigned values and assigned intervals for different control specimens: 3 laboratories, 15 duplicate determinations each (3 × 15) and 5 laboratories, 5 duplicate determinations each (5 × 5). Values for chloride are the same in mmol/l as in mval/l.

Constituent (unit)	Protocol	Assigned values (assigned intervals)				
Method		Reg. No. 19500	19600	19700	22500	22600
Chloride (mval/l) Coulometry	3 × 15	94 (90–98)	93.5 (90–97)	92.5 (89–96)	104 (100–108)	113 (109–117)
	5 × 5	94 (91–97)	93.5 (90–97)	92 (89–95)	103 (100–106)	110.5 (106–115)
Glucose (mg/100 ml) Hexokinase reaction	3 × 15	107 (101–113)	109 (102–116)	99 (93–105)	89.5 (83–96)	209 (196–222)
	5 × 5	107 (100–114)	109 (101–117)	97 (90–104)	93 (83–101)	214 (203–225)
α-Hydroxybutyrate dehydrogenase (U/l)	3 × 15	151 (141–161)	134.5 (125–144)	136 (125–147)	129 (110–148)	153 (133–173)
Standard method	5 × 5	150 (135–165)	133 (122–144)	142 (129–155)	132 (109–155)	160 (133–187)
Alkaline phosphatase (U/l) Standard method	3 × 15	198 (172–224)	213 (186–240)	205 (179–231)	78 (68–88)	205 (180–230)
	5 × 5	206 (172–240)	216 (186–246)	216 (192–240)	79 (70–88)	209 (186–232)

The amount of dispersion is an important factor in the selection of a control specimen for internal control of accuracy for a given constituent-method combination; in some cases the imprecision is so large that the specimen is of little use in accuracy control.

Such strict requirements for the selection of laboratories to serve as reference laboratories and for the adherence to carefully defined conditions during the analytical process mean, of course, that the number of possible reference laboratories is limited.

It must be kept in mind, however, that quality control according to the Guidelines of the MSWG, including the control specimens used in this process, is intended to provide an ongoing control of the whole process of clinical laboratory analysis from the reliability and appropriate use of volumetric equipment, standards, reagents and measuring apparatus to the calculations and the basis on which they are made. This is possible only if the analytical results have been obtained under carefully defined conditions (analytical procedure and protocol) and the assigned values and assigned intervals have been selected in a manner appropriate to the structure of the data. Other advantages of having carefully defined conditions, including the information that can then be obtained, are discussed in Section 4.5.

An additional reason for requiring carefully defined conditions at least in the Federal Republic of Germany is that here quality control according to the Guidelines of the MSWG is a legally approved alternative to official calibration. Thus court cases contesting the reliability of assigned values and assigned intervals can be expected. A further reason for expecting such disputes is that some of the bodies regulating payments to physicians for health care (Kassenärztliche Vereinigungen) have decided not to pay for certain clinical laboratory tests unless the physician performing the test has valid certificates indicating successful participation in the appropriate interlaboratory surveys.

In such disputes assigned values can be shown to be reliable only if the analytical results on which they are based have been obtained under the defined, reproducible conditions mentioned above.

4.2 Requirements to be Met by the Protocol

In each reference laboratory enough analytical results per constituent-method combination must be obtained under carefully defined conditions, so that it is possible to provide the following information for each combination used in a particular laboratory:

1. a frequency distribution for the analytical results,
2. a location parameter for the analytical results,
3. a measure of day-to-day imprecision,
4. a measure of imprecision in the series.

It should be possible to use these parameters to compare the results from the different reference laboratories. If there are marked differences in the day-to-day imprecision, it is often possible to obtain an indication of possible sources of error from the imprecision in the series.

Such an experimental design must also be practicable, and the procedure for selecting assigned values must take into consideration the characteristics that the above-mentioned parameters are found to have.

4.3 Results of a Comparison of the Parameters Characterizing the Analytical Results from Different Laboratories

A comparison of the distributions of results from different laboratories showed that for a certain number of constituent-method combinations the pooled results from different laboratories are not a homogeneous sample; rather, they are a conglomerate containing samples with distributions that are sometimes quite different (Section 3.4 and Table 3-7).

A comparison of the standard deviations of the analytical results from different laboratories for the same constituent-method combination frequently yielded significant differences. Standard deviations do not even have a characteristic size for a particular laboratory but can differ from specimen to specimen, as was found when different specimens were analyzed in the same runs (Section 3.5 and Tables 3-8 and 3-9).

As a result, no predictions can be made about the relative location and amount of dispersion of sets of results from different laboratories. Even for relatively specific methods the values of these parameters are dependent on the characteristics of the matrix of the specimen in question. This means that for each specimen these parameters must be determined for each constituent-method combination in each laboratory and examined prior to assigned value determination.

It should be noted that in our evaluation we eliminated neither distributions with a location quite different from the others nor distributions with a large dispersion. Rather, we tried through inquiry and, where necessary, repetition of analyses or other investigations to find explanations for such phenomena (Section 3.6.3). This enabled us to obtain important information on how to improve standards, methods and control specimens (Section 4.5), information that would have been lost in a protocol where statistical outliers are automatically eliminated.

4.4 The Protocol

The findings in our examination of the location and dispersion parameters of the analytical results from different reference laboratories demonstrate that it is not possible to assume a common probability distribution let alone a normal distribution; rather, we recommend that a P% interval for the analytical results be determined without any assumptions being made about the distribution and that a suitable location parameter (e. g. the midpoint) be selected for this interval and designated the assigned value.

The reasons why we specify duplicate determinations on different days with 15 days per reference laboratory for lyophilized control specimens and 10 for liquid control specimens were discussed in Section 2.3.

In deciding how many laboratories should participate in an assigned value determination per constituent-method combination the following limiting factors must be considered:

1. the volume of the homogeneous lot of a control specimen,

2. the amount of money available for the analyses in the reference laboratories and how this affects the selling price of the control specimen,

3. the time that elapses between the availability of the control specimen for analysis and the final establishment of the assigned values. This should be as short as possible. The period in which the control specimen can be marketed is reduced by this time span, and interest must be paid on the money needed for material and for preparation of specimens.

This leads to the requirement of designing an optimal protocol, in which the experimental conditions are defined carefully enough, the number of determinations is large enough and the number of reference laboratories is so chosen that after completion of the determinations the most that will usually be necessary is a few inquiries to the participants, with as few subsequent investigations as possible, for these would require a disproportionate amount of time.

In the course of assigned value determinations in over 200 lots it could be established that three reference laboratories per constituent-method combination is usually adequate. In a very small number of cases, where it was found necessary to check for possible differences in analytical results obtained with reagent kits from different manufacturers (e. g. cholinesterase and acid phosphatase), the number was increased.

A comparison of the protocol using 3 X 15 duplicate determinations with that using 5 X 5 duplicate determinations as applied to five different control specimens from two manufacturers (Table 3-15) showed good agreement in the values found for the location parameter, but for some constituent-method combinations there were differences in the amount of dispersion. In order to determine the cause of these differences it was necessary to conduct additional investigations in several laboratories in the 5 X 5 study because of the small amount of data. As a result the assigned value determination took at least as long here as with the 3 X 5 design.

With small numbers of results per laboratory, deviations of the location parameter and the dispersion of a particular laboratory are not so obvious and thus may not lead to the desired identification of causes.

4.5 Advantages of the Protocol

Because of the carefully defined conditions under which the analyses are carried out, which lead to good com-

parability, and because of the number of determinations required per laboratory, the protocol often permits identification of the causes of differences (a) in the results from different reference laboratories and (b) between the assigned value and the mean of the participants in the interlaboratory surveys where the control specimens are used.

The identification of these causes has led to improvements in the characteristics of control specimens, standards, methods including the reagents, and equipment and in their use. One or two examples for each of these cases are given below.

4.5.1 Control specimens

For a good number of constituents it was found that in order to insure effective monitoring, special safety precautions had to be introduced, or changes made in the manufacturing process or in the instructions for use.

Turbidity: After reconstitution, lyophilized control specimens may be much more turbid than native human serum, depending on the original material and how lyophilization was carried out. This can interfere markedly with the analysis of certain constituents (e. g. total protein, serum iron).

The Reference Commission discussed this problem, which in some cases is method-dependent, with the manufacturers of control specimens; thereafter the manufacturers took steps which substantially reduced this interference factor or eliminated it completely.

Potassium: After part of a lot of a liquid control specimen was packaged a significant and reproducible drop in the potassium concentration was observed. This could be attributed to a new supply of glass containers.

Glucose: In lyophilized control specimens, a measurable initial drop in the glucose concentration is sometimes observed that is dependent on characteristics of the matrix and on storage conditions and time; this is referred to as ripening. Such changes in concentration may also be method-dependent. Based on information from assigned value determinations, interlaboratory surveys and various post-analysis investigations, the manufacturers have now standardized the conditions for ripening to such an extent that the user is no longer aware of any changes in concentration.

Bilirubin: Bilirubin in control specimens is sometimes more sensitive to light than it is in native serum specimens. The observation of this phenomenon during assigned value determination led to a more careful study of the problem and subsequently to appropriate instructions to the user.

Creatine kinase: Analogous observations have been made recently for the measurement of the activity of this enzyme.

Alkaline phosphatase: There is an increase in the activity of alkaline phosphatase following reconstitution, which often reaches a plateau. The amount of increase and the time it takes until the plateau is reached can vary from lot to lot of a control specimen. The level of the plateau may be influenced by the temperature at which reconstitution took place. A discussion of the differences in the location and dispersion of the results from different laboratories observed during assigned value determination has led to a more careful study of this phenomenon in each lot, to a standardization of conditions for use of this enzyme and to the rejection of unsuitable enzyme preparations.

4.5.2 Standards

In a few cases the differences between the analytical results from different reference laboratories could be attributed to the standards used or the method of calculating the results.

Two examples are:

Protein determination: Relatively large variations within and between the reference laboratories were reduced substantially (24) by using samples of the same bovine albumin standard which conformed in almost all of its properties to the specification of *Peters* (22, 23). In the meantime this standard has come into general use for the determination of total protein.

Glucose determination: When glucose was determined with the hexokinase reaction (25), systematic differences were found between the reference laboratories that were dependent on the method of calculation used (primary standard or molar extinction coefficient). This led to a reevaluation of the molar extinction coefficients for NADPH and NADH and a subsequent correction (26). A manufacturer who had been testing and adjusting his standards for other methods of glucose determination with the hexokinase reaction thereafter switched to primary standards.

4.5.3 Methods and reagents

Emission flame photometry: Significant differences were found between photometers with and without the lithium guideline, especially for sodium determination. As a result, the manufacturer of one photometer without a lithium guideline for the diluent added lithium (27); thereupon significant differences between this photometer and those with lithium guideline were no longer found.

Serum iron determination: Several years ago in interlaboratory surveys, marked differences were found repeatedly between the assigned values of the reference laboratories and the means of the participants, the extent depending on which reagent kits had been used.

Although the reference laboratories have not changed their methods, these differences have not been seen during the past two years. Thus the change can be attributed to an improvement in the reagent kits.

Enzyme activity determinations: In internal control of accuracy and in assigned value determinations, results which had been obtained with reagents from different manufacturers showed such great differences that effective accuracy control was impossible.

The German Society for Clinical Chemistry then formed an enzyme commission with instructions to determine the causes of these differences and develop recommendations for standardized conditions for the determination of enzyme activities. Members of this commission were scientists from large hospital laboratories, research laboratories and industry with experience in this field. Their efforts resulted in the Recommendations of the German Society for Clinical Chemistry (28–31), which are now used all over the Federal Republic for more than 90% of all enzyme activity determinations.

The comparison of results from reference laboratories with each other and with results from participants in interlaboratory surveys continues to lead to new information about important but previously unknown influences and interference factors and subsequently to their standardization or elimination.

4.6 Comparison with Other Protocols

A number of approaches have been suggested for the determination of assigned values and decision limits for interlaboratory surveys. The advantages and disadvantages of these approaches, as compared with the method we have just described and which we have now been using for a good number of years, must be included here.

4.6.1 Consensus value

Whitehead (32) and subsequently *Jansen, van Kampen* and coworkers (33) have developed a procedure for the determination of decision limits for interlaboratory surveys that is based on the mean of the results of the participants in the interlaboratory survey. Here the mean and standard deviation are calculated from the results of single or duplicate determinations by all participants who used the same analytical principle. The large number of participants with only a few results has the effect that the conditions for analysis are not clearly defined and that the reliability of the results from the individual laboratories cannot be evaluated.

It is true that for unproblematical analytical methods the location parameters are in good agreement, as we could show in several comparative studies on the same control specimens. But if differences arise between the consensus value and our assigned value, no information that might be helpful in identifying the causes can be provided by the protocol with unclearly specified conditions for analysis.

The determination of the dispersion parameter after truncation is arbitrary. It is perfectly conceivable that this truncation can lead to the elimination of the results from laboratories with especially good reliability (e. g. enzyme activity determinations).

The analytical results and the location and dispersion parameters calculated from them are obtained under conditions that are not clearly defined. Therefore they cannot serve as an alternative to official calibration or as the basis for a legal decision.

4.6.2 The use of fixed decision limits

Because of difficulties in defining decision limits on the basis of the dispersion of the analytical results of interlaboratory survey participants, the suggestion has been made that the allowable deviation from the assigned value for internal control of accuracy and for interlaboratory surveys be fixed independent of the particular control specimen. According to this suggestion, such fixed decision limits are to be selected, for example, on the basis of the results of interlaboratory surveys with a special group of laboratories by calculating the mean of the standard deviations obtained over a long period of time.

It could be shown (see Section 3.5) that the amount of dispersion depends, among other things, on the specimen used (matrix characteristics). As a result, such fixed decision limits would be too narrow for some control specimens and too wide for others and thus clearly unsuitable. Furthermore, control specimens that are unsuitable for monitoring a constituent in internal accuracy control or in interlaboratory surveys would not be identified on the basis of their dispersion parameters and thus would not be rejected.

The fixing of allowable deviations on the basis of maximum allowable imprecision from day to day according to the Guidelines of the MSWG (11, 12), i. e. 5% (in exceptional cases 10%) results in a range that is certainly too wide if one considers diagnostic requirements and the level of precision actually obtainable for many constituents; these limits should be considered as "penalty limits" and should not be regarded as either desirable or necessary in clinical use.

In their suggestions *Ludewigs, Rotzler & Völkert* (34) do not take into consideration any of the observations on data structure made above. These authors suggest that the well-equipped reference laboratories under highly qualified direction selected for assigned value determination be replaced by a large number of first class interlaboratory survey participants. The conditions under which analyses are made cannot be defined as exactly with this large number of laboratories as in the reference laboratories. Beyond this, the usefulness of a location parameter of an individual laboratory depends

on its reliability. The authors have succumbed to the fascination of large numbers of participating laboratories; a majority opinion takes the place of defined conditions. In some cases, the results will still be comparable, even though they may no longer be accurate. But accuracy is vital for the evaluation of clinical laboratory results by means of comparison with normal ranges.

5. Conclusions

On the basis of observations made in the course of many assigned value determinations, the following conclusions can be drawn:

1. In many quantitative clinical chemical analyses the analytical results contain nonspecific components of various sizes whose kind and quantity depend on the matrix of the particular specimen.

2. For a great many reagents used in clinical chemistry there are changes during the period of use of a particular lot and differences between lots that can have an effect on the analytical results.

3. Thus for many constituents and methods it is not appropriate to carry out internal or external accuracy control by comparing the analytical result for the accuracy control specimen with the best estimate (by measurement) of the "true value" of the relevant constituent. In this approach important components of the result (e. g. influence of the matrix and the reagents) are not taken into consideration.

4. Therefore accuracy control of routine methods in clinical chemistry with control specimens must be carried out by comparing analytical results with values termed assigned values and the related assigned intervals.

5. These assigned values must be determined on the basis of analyses in routine runs carried out under carefully specified conditions in a number of specially equipped laboratories, whose heads are particularly highly qualified.

6. The experimental design must be such that it is possible to obtain reproducible and sufficiently reliable information on the structure of the data, the amount of dispersion and the location for each participating reference laboratory.

7. From the data obtained in assigned value determinations in over 200 lots of control specimens (three reference laboratories usually participated for each constituent-method combination, and each reference laboratory carried out duplicate determinations on 10 successive working days for liquid control specimens or 15 successive working days for lyophilized specimens) the following was found:

7.1 If in a single reference laboratory parallel analysis of different lots of the same control specimen is carried out in the same runs and according to the same protocol, both the within-run imprecision and the day-to-day imprecision can show significant differences. The greatest standard deviation is often as much as three times the smallest standard deviation.

7.2 If samples of the same control specimen are analyzed in different reference laboratories, the results are often so different with regard to their distribution that they cannot be considered to be homogeneous. Of special importance, the concept of the same precision in all laboratories usually proves to be inapplicable.

8. As a result, the assigned intervals and assigned values are selected without any special assumptions being made about the distribution. The results from each reference laboratory are evaluated separately and then the results from all laboratories are pooled. A suitable interval that includes at least 95% of the results is then selected as the assigned interval. With a view to the Guidelines of the MSWG, the midpoint is selected as the location parameter of this interval. For reasons of stability, other location parameters such as the median would be preferred.

9. The following practical points should be considered:

The choice of three reference laboratories per constituent-method combination is a practicable minimum.

At least 10 duplicate determinations should be carried out by each reference laboratory.

An assigned value should be based on no fewer than 50 analytical results.

10. Determination of assigned values according to carefully specified analytical and statistical procedures has distinct advantages. This process of assigned value determination permits identification of the reasons for deviations among the reference laboratories. When assigned values are used, the reasons why individual values deviate in individual laboratories or the mean of the participants deviates from the assigned value in interlaboratory surveys can frequently be determined. In the past, the identification of such factors has led to marked improvements in the characteristics of control specimens, standards and methods (including reagents and equipment) and in how they are used.

References

1. Broughton, P. M. G. & Raine, D. N. (1969), Quality control in clinical biochemistry. Ann. Clin. Biochem. *6*, 89–148.
2. Büttner, H. (1967), Statistische Qualitätskontrolle in der Klinischen Chemie. Z. Klin. Chem. Klin. Biochem. *5*, 137–145.
3a. Büttner, H., Hansert, E. & Stamm, D. (1974), Auswertung, Kontrolle und Beurteilung von Meßergebnissen. In: Methoden der enzymatischen Analyse, 3rd ed. (Bergmeyer, H. U., ed.) Verlag Chemie, Weinheim/Bergstraße, Vol. 1, pp. 339–422.

3b. Büttner, H., Hansert, E. & Stamm, D. (1974), Statistical analysis, control and assessment of experimental results. In: Methods of Enzymatic Analysis, 2nd ed. (Bergmeyer, H. U., ed.) Academic Press, New York, London, Vol. 1, pp. 318–395.
4. Büttner, J., Borth, R., Boutwell, J. H., Broughton, P. M. G. & Bowyer, R. C. (1977), Provisional recommendation on quality control in clinical chemistry, Part 3. Calibration and control materials. Clin. Chim. Acta *75*, F 11–20, J. Clin. Chem. Clin. Biochem. *15*, 233–238.
5. Büttner, J., Borth, R., Broughton, P. M. G. & Bowyer, R. C., Provisional recommendation on quality control in clinical chemistry, Part 4. Internal quality control. Clin. Chim. Acta (in press), J. Clin. Chem. Clin. Biochem. (in press).
6. Stamm, D. (1972), Qualitätskontrolle klinisch-chemischer Analysen. Ein Lernprogramm in 5 Lektionen für medizinisch-technische Assistentinnen, Studenten und Ärzte. Programmiert von Zielinski, J., Tulodziecki, G. und Mitarbeitern. Georg-Thieme-Verlag, Stuttgart.
7. Stamm, D. (1974), Calibration and quality control materials. Z. Klin. Chem. Klin. Biochem. *12*, 137–145.
8. Stamm, D. (1979), Reference materials and reference methods in clinical chemistry. J. Clin. Chem. Clin. Biochem. *17*, 283–297.
9. Gesetz über das Meß- und Eichwesen (Eichgesetz) vom 11. Juli 1969, Bundesgesetzblatt vom 15. Juli 1969, Part I, 759–770.
10. Verordnungen über Ausnahmen von der Eichpflicht (Eichpflicht-Ausnahmeverordnung) vom 26. Juni 1970, Bundesgesetzblatt vom 30. Juni 1970, Part II, 960–965.
11. Richtlinien der Bundesärztekammer zur Durchführung von Maßnahmen der statistischen Qualitätskontrolle und von Ringversuchen im Bereich der Heilkunde. Deutsches Ärzteblatt *68*, 2228–2231 (1971).
12. Ausführungsbestimmungen und Erläuterungen zu den Richtlinien der Bundesärztekammer zur Durchführung der statistischen Qualitätskontrolle und von Ringversuchen im Bereich der Heilkunde. Deutsches Ärzteblatt, *71*, 961–965 (1974).
13. Stamm, D. & Büttner, H. (1969), Ringversuch zur Qualitätskontrolle 1968. Z. Klin. Chem. Klin. Biochem. *7*, 393–403.
14. Stamm, D. (1971), Ringversuche in der Klinischen Chemie. Schweiz. Med. Wochenschr. *101*, 429–437.
15. Stamm, D. (1975), The determination of assigned values for control specimens. In: Quality Control in Clinical Chemistry (Anido, G., van Kampen, E. J., Rosalki, S. B. & Rubin, M., eds.) Walter de Gruyter, Berlin, New York, pp. 117–130.
16. The SI for the health professions. World Health Organization. Geneva 1977.
17. International Organization for Standardization: Precision of test methods––Determination of repeatability and reproducibility. Draft International Standard ISO/DIS 5725, Oct. 1977.
18. Pearson, E. S. & Hartley, H. O. (eds.) (1970), Biometrika Tables for Statisticians, Vol. I. Cambridge University Press, Cambridge, 3rd ed., reprinted with additions.
19. Cochran, W. G. (1941), The distribution of the largest of a set of estimated variances as a fraction of their total. Ann. Eugenics *11*, 47–52.
20. Huber, P. I. (1972), Robust statistics––a review. Ann. Math. Stat. *43*, 1041–1067.
21. Grubbs, F. E. (1969), Procedures of detecting outlying observations in samples. Technometrics *11*, 1–21.
22. Peters, T. Jr. (1968), Proposals for standardization of total protein assays. Clin. Chem. *14*, 1147–1159.
23. National Committee for Clinical Laboratory Standards: Total serum protein (bovine serum albumin). Approved Standard ASC-1, 1973. 771 E. Lancaster Avenue, Villanova, Pa. 19085.
24. Müller, B. & Stamm, D. (1973), Experiences with a protein standard in clinical chemistry. In: Organisation des Laboratoires––Biology Prospective (Siest, G., ed.) L'Expansion Scientifique Francaise, Paris, pp. 543–546.
25. Bergmeyer, H. U., Bernt, E., Schmidt, F. & Stork, H. (1974), D-Glucose, Bestimmung mit Hexokinase und Glucose-6-phosphat-Dehydrogenase. In: Ref. 3a, Vol. 2, pp. 1241–1250 and Ref. 3b, Vol. 3, pp. 1196–1201.
26. Bergmeyer, H. U. (1975), Neue Werte für die molaren Extinktionskoeffizienten von NADH und NADPH zum Gebrauch im Routine-Laboratorium. Z. Klin. Chem. Klin. Biochem. *13*, 507–508.
27. Eppendorf-Flammenphotometer: Arbeitsanleitung V 734, 1974, Eppendorf Gerätebau, Barkhausenweg 1, D-2000 Hamburg 63.
28. Empfehlungen der Deutschen Gesellschaft für Klinische Chemie, Standardisierung von Methoden zur Bestimmung von Enzymaktivitäten in biologischen Flüssigkeiten. Z. Klin. Chem. Klin. Biochem. *8*, 658–659 (English translation pp. 659–660) (1970).
29. Empfehlungen der Deutschen Gesellschaft für Klinische Chemie, Standardisierung von Methoden zur Bestimmung von Enzymaktivitäten in biologischen Flüssigkeiten. Experimentelle Begründung der optimierten Standard-Bedingungen. Z. Klin. Chem. Klin. Biochem. *10*, 182–192 (English translation pp. 281–291) (1972).
30. Empfehlungen der Deutschen Gesellschaft für Klinische Chemie, Standardisierung von Methoden zur Bestimmung von Enzymaktivitäten in biologischen Flüssigkeiten. Experimentelle Begründung der optimierten Standard-Bedingungen, Ergänzungen zur Standard-Methode zur Bestimmung der Aktivität der Glutamat-Dehydrogenase (GLDH). Z. Klin. Chem. Klin. Biochem. *12*, 391 (English translation p. 392) (1974).
31. Empfehlungen der Deutschen Gesellschaft für Klinische Chemie, Standardisierung von Methoden zur Bestimmung von Enzymaktivitäten in biologischen Flüssigkeiten, Standard-Methode zur Bestimmung der Aktivität der Creatin-Kinase. Überarbeitete Fassung von 1976. J. Clin. Chem. Clin. Biochem. *15*, 249–254 (English translation pp. 255–260) (1977).
32. Whitehead, T. P., Browning, D. M. & Gregory, A. (1973), A comparative survey of the results of analyses of blood serum in clinical chemistry laboratories in the United Kingdom. J. Clin. Pathol. *26*, 435–445.
33. Jansen, A. P., van Kampen, E. J., Leijnse, B. Meijers, C. A. M. & van Munster, P. J. J. (1977), Experience in the Netherlands with an external quality control and scoring system for clinical chemistry laboratories. Clin. Chim. Acta *74*, 191–201.
34. Ludewigs, M., Rotzler, A. & Völkert, E. (1978), Bewertung von Ringversuchsergebnissen. Diagnostik *11*, 163–165.
35. Albert, G. E. & Johnson, R. B. (1951), On the estimation of central intervals which contain assigned proportions of a normal univariate population. Ann. Math. Stat. *22*, 596–599.

Prof. Dr. Dr. D. Stamm
Dr. E. Hansert
Max Planck Institut für Psychiatrie
Kraepelinstraße 10
D-8000 München 40